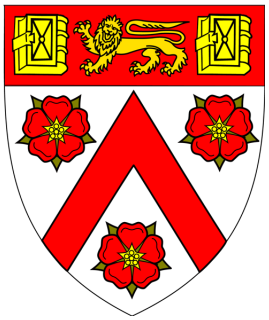
The Impact of Randomisation in Load Balancing and Random Walks



Leran Cai
Trinity College

A dissertation submitted to the University of Cambridge in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Computer Science

University of Cambridge
Computer Laboratory
William Gates Building
15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom

Email: lc647@cl.cam.ac.uk

August 18, 2021

Declaration

I, Leran Cai of Trinity College, being a candidate for the Doctor of Philosophy in Computer Science, hereby declare that this thesis – *The Impact of Randomisation in Load Balancing and Random Walks* – and the work described in it are my own work, unaided except as may be specified below, and that the thesis does not contain material that has already been used to any substantial extent for a comparable purpose.

Signed: *Leran Cai*

Date: 18th Aug 2021

Abstract

The real world is full of uncertainties. Classical analyses usually favour deterministic cases, which in practice can be too restricted. Hence it motivates us to add in *randomness* to make models similar to practical situations. In this thesis, we mainly study two network problems taken from the distributed computing world: iterative load balancing and random walks. An interesting observation is that the problems we study, though not quite related regarding their real world applications, can be linked by the same mathematical toolkit: Markov chain theory. These problems have been heavily studied in the literature. However, their assumptions are mostly *deterministic*, which causes less flexibility and generality to the real world settings. The novelty of this thesis is that we add randomness in these problems in order to observe worst cases vs. average cases (load balancing) and static cases vs. dynamic cases (random walks).

For iterative load balancing, the randomness is added on the number of tasks over the entire network. Previous works often assumed worst case initial loads, which may be wasteful sometimes. Hence we relax this condition and assume the loads are drawn from different probability distributions.

In particular, we no longer assume the initial loads are chosen by an adversary. Instead, we assume the initial loads on each processor are sampled from independent and identically distributed (i.i.d.) probability distributions. We then study the same problems as in classical settings, i.e., the time needed for the load balancing process to reach a sufficiently small discrepancy.

Our main result implies that under such a regime, the time required to balance a network can be much faster. An insightful observation is that the load discrepancy is proportional to the term $t^{-1/4}$ where t is the time used to run the protocol. This implies two main improvements compared with previous works: first, when the initial discrepancy is the same, our regime can reach small discrepancy faster; second, we have established a connection between the time and the discrepancy while previous analyses do not have.

For random walks, the randomness is added on the network topologies. This means at each time step (considering discrete times), the underlying network can change randomly. In particular, we want the graph "evolves" instead of changing arbitrarily. To model the graph changing process, we adopt a model commonly used in the literature, i.e., the edge-Markovian model. If an edge does not exist between the two nodes, then it will appear in the next step with probability p, and if it does then in the next step it will disappear with probability q. This model can simulate real world scenarios such as adding friends with each other in social networks or a disruption between two remotely connected computers.

Our main contributions regarding random walks include the following results. First, we divided the edge-Markovian graph model into different regimes in a parameterised way. This provides an intuitive path to similar analyses of dynamic graph models. Dynamic models are often hard to analyse in the field because of its complicated nature. We present a possible strategy to reach some feasible solutions by using parameters (p, q above) to control the process. Second, we again analyse the random walk behaviours on such models. We have found that under certain regimes, the random walk still shows similar behaviours especially its mixing nature as in static settings. For the other regimes, we also show either weaker mixing or no mixing results.

Acknowledgement

Undertaking this PhD has been a life-changing experience for me and it would not have been possible to be achieved without the support and guidance from many people.

I would like to say a big thank you for the help from my supervisor Thomas Sauerwald and my group colleagues: Luca Zanetti, Hayk Saribekyan, Nicolas Rivera and John Sylvester. Their deep insightful understandings in mathematics made my PhD life very enjoyable and I learned a lot from them. It's good to have a bunch of brilliant and smart people around me.

I gratefully acknowledge the funding from Trinity college. Without Trinity, I would not have been able to start this PhD life. Trinity gave me my second big chance in my not so successful life at my 23rd year.

Thanks to Zhenguo Li, Yitong Sun and Jiacheng Sun who offered me an incredible opportunity to work as an intern at Huawei's research group. So I can not only make a living but also see how the mathematics I learned during my PhD can be used further.

I appreciate the intriguing discussions with Yuchen Yang and Haoyang Liu during my PhD time.

Thank my wife Xin Zhou for her love and company during this hard time. I hope her name will appear everywhere in my future life.

Many thanks to my mother for her love and endless support for my life. May her rest in peace.

Contents

List of Figures

List of Tables

List of Notations

$\mathbf{P}, \mathbf{M}$	Markov chain transition matrices
$\mathbb{E}$	The expectation of a random variable
$\mathbb{P}$	The probability of a event
$\lambda(\mathbf{P})$	The second largest eigenvalue of the matrix $\mathbf{P}$
log	Logarithm, unless stated otherwise in this thesis the base is e
μ, ν	Probability distributions
$\|\cdot\|_{TV}$	The total variation distance of two probability distribution
Uni	Uniform distribution
Bin	Binomial distribution
Poi	Poisson distribution
Geo	Geometric distribution

Chapter 1

Introduction

1.1 Problems and motivation

1.1.1 Distributed computing and our problems

Distributed computing is a field of computer science which studies networks of computers that communicate with each other by sending and receiving messages. Nowadays, networks are everywhere due to the widespread of the Internet and developed infrastructure. Networks not only connect the world but also provide us with high-performance computing models because a big computing task can be solved on a distributed system efficiently while it may take years to finish on a single computer.

The idea of distributed computing can be seen in many situations where a large number of queries and short response times are required. For example, massively multiplayer online games, peer-to-peer applications, and online shopping providers like Amazon. It can also be seen in situations where the number of related parties is massive, e.g., social networks like Facebook and Twitter, citation networks, web graphs, etc. Usually, real world networks may include the following features:

1. Small average degree
2. Small diameter
3. High connectivity and expansion
4. High clustering coefficients, especially for social networks
5. Dynamic behaviour

A computer in a network, even a very large network, may only communicate with its neighbours because it has a connection limit, and also communicating with every other computer in the network is wasteful in most cases. The diameter of the network is

the maximum distance between any pair of computers. Usually, a computer needs to communicate with another one in the network even if they are not directly connected. Hence a real-world network tends to have a small diameter to guarantee efficiency. High connectivity or expansion means that the network should remain connected with high probability when some edges are removed, which corresponds to possible interruptions between computers in a real network. Some networks like social networks are likely to have clusters, which means among some computers the connections are significantly more than those among these computers and the rest of the network. Real-world networks are facing accidents every day. For instance, existing nodes or edges may disappear in a network and new nodes or edges may be added.

Due to these features, it is important to design efficient local algorithms, i.e., algorithms with the restrictions that any computer does not hold global information to make decisions or strategies, to save resources while still achieving high-demanding tasks. Hence it is natural to consider balancing the workload over the entire network, which leads to the first topic of this thesis: **load balancing**. In particular, this means that each processor only communicates with its neighbours and passes tasks from a high workload machine to a lower workload machine.

The second topic of this thesis is **random walks**. They do not seem to have a direct connection with distributed computing. They serve more as an integral subroutine in distributed computing, such as load balancing [52], searching [29], pageranking [24, 82, 38], voting [30, 31], and information propagation [53, 21]. In this thesis, we study random walks on graphs. A random walk on a graph typically means that a particle jumps among the nodes in the graph. Each time the particle picks one of its neighbours and jumps to it with some well-defined probability [60, 31, 33].

1.1.2 Motivation

1.1.2.1 Motivation of using Markov chain theory

Although load balancing problems and random walks look quite irrelevant to each other, the studies of them can be linked by **Markov chain** theory since both of them can be modeled by it. We delay the formalisation in later chapters but give a plain explanation in this section.

A Markov chain is a random process $X_0, X_1, ...$ where X_is are chosen from a set of states Ω. The indices often represent time steps in the process. From t to $t+1$, the process jumps from the state X_t to X_{t+1} with a probability p. For any two states $u, v \in \Omega$, the transition probability from u to v is set as a fixed value $p_{u,v}$. In particular, if $X_t = u$, then with probability $p_{u,v}$, the process will jump from the state u to $X_{t+1} = v$. If Ω is finite, then for each X_i, it has a corresponding probability distribution μ_i such that $\mu_i(u)$ is the probability that $X_i = u$.

With the above quick introduction on Markov chains, we can now understand why we can link load balancing problems and random walks by it. Assuming our distributed computing problem is defined on a network $G = (V, E)$ where V is the node set and E is the edge set. Regarding load balancing, we can use μ_i to represent the load distribution at time i. Simply normalising it would make it look like a probability distribution. With respect to random walks, the probability $\mu_i(u)$ is exactly the probability that the particle is at the node u at time i. Hence for both problems with standard settings, they can be transferred to Markov chains.

In addition, Markov chain theory has been intensively studied in the past decades [59]. Its applications in distributed computing have been widely used in the field, e.g., [75, 41]. We, therefore, choose to use it as our main toolkit because it has been proven to be suitable for our analysis. Also, we would like to contribute more to this area by adding more realistic assumptions to the theoretical problems.

1.1.2.2 Motivation of using randomness

As real-world events are more or less unpredictable, it motivates us to add a bit of *randomness* when analyzing and building our mathematical models. Load balancing and random walks have been mainly studied under fixed settings and regimes. In other words, classical settings tend to have static graphs and strict quantities.

Real-world networks, however, tend to be *large* and *complex*. This implies that a global description is impossible. Hence researchers turn to *local* descriptions, such as by which local rules vertices are connected, or what the maximum degree is. The local rules are preferred to be *random* because it can include various situations inside a huge network by a few simple statements. In other words, randomness is an efficient way to deal with the complexity of the network. Though it may not be able to describe a large and complex network precisely, it decreases the difficulty of analysing such a network while a global description is unnecessarily complicated and also intractable. In the book by Newman [68], there are many empirical properties of and scientific methods for networks where random graphs play a very important role. They also support the argument that using randomness is a very efficient way to deal with large complex networks.

With respect to the above concerns, we consider more randomness in our regimes to make our models approach real-world scenarios. For load balancing problems, we assume randomness on the number of tasks distributed on the network. For random walks, we assume dynamic evolving underlying networks.

We now introduce how to add randomness to these problems. Intuitively, if we want to add the randomness somewhere, it can either be on the nodes or the edges of a network. If we add the randomness on the nodes, then making the loads on each node random looks like a natural choice. If the randomness is put on the edges, then it leads us to think of the possibility that the edge can be broken or created. Hence we have the following two

research directions: 1. load balancing with random loads; 2. random walks on evolving graphs.

1.2 Load balancing with random input loads

Load balancing, as previously introduced, means we balance the load distributed over a network. Given a large network, which is usually a group of connected processors, each processor handles a number of tasks which we call *tokens*. In practice, it is natural for us to expect that the load distribution is as even as possible in order to maximize the utility of all processors. Since the real network is huge and each processor can only communicate with its neighbours, a global solution is hard to achieve because that usually requires a synchronisation among all nodes, which is very wasteful. Hence, we usually prefer local protocols to deal with such problems. Therefore one important and classic problem is the following: if we know the initial difference between the maximum load and the minimum load is K, what algorithm can we use to reduce the difference by only allowing the processors to transfer tokens locally and how small can that difference become?

Previous works always consider K directly or pick an initial load distribution by using an adversary. However, statistically speaking, we can usually assume there exists a hidden distribution or a function model for the real-world data. We believe that it would be more realistic to assume that the initial token distribution is distributed according to some distributions. This is where we add randomness for load balancing problems. In other words, we do not assume the initial maximum difference is K, but instead assume that the number of tokens at each node is sampled from a distribution with a range equal (or close) to K. Since more assumptions are added, we will explore the previous algorithms again and also expect that the results can be improved. Adding this new randomness would change the core of the original problem. Hence this is novel and worth studying.

1.3 Random walks on evolving graphs

A random walk is a random process where a particle starts from a vertex in a graph and at each step the particle jumps to a randomly chosen neighbour. It is also a special case of a Markov chain.

Previous works mainly studied static graphs, e.g., [53, 21], which means that the underlying graph that the particle moves on does not change over time. However, this is not always true. Real networks tend to evolve, which changes the behaviour of random walks on networks significantly. Therefore we are interested in studying random walks on evolving graphs and investigating what their behaviour in such an environment is. Dynamic graph models turn out to be very hard to analyze. In order to handle it, we try to control the dynamic of the model by some parameters in this thesis. Our analysis

can be potentially used elsewhere as long as the other dynamics create random graphs instead of following some deterministic rules.

1.4 Related works and our main contributions

In this introductory chapter, we only present the most related works in the field to emphasize how we improve the key results. More detailed literature reviews will be provided in Chapter 3 and 4.

Regarding **load balancing** problems, to the best of our knowledge, before our work, the previous works mainly focused on fixed settings. The main question is concerned about how fast the protocol can reach a small *discrepancy*, the difference between the maximum and the minimum load. Classical works include [75, 41, 66, 85]. We compare our results with some of the classical ones below:

Rounds	**Discrepancy**	**Reference**
$O\left(\frac{\log(Kn)}{1-\lambda}\right)$	$O\left(\frac{d\log n}{1-\lambda}\right)$	Rabani, Sinclair, and Wanka (FOCS'98)
$O\left(\frac{d\cdot\log(Kn)}{1-\lambda}\right)$	$O\left(\min\left\{\sqrt{\frac{d\log n}{1-\lambda}}, \frac{d\log\log n}{1-\lambda}\right\}\right)$	Friedrich and Sauerwald (STOC'09)
$O\left(\frac{\log(Kn)}{1-\lambda}\right)$	$O(1)$	Sun and Sauerwald (FOCS'12)
$O\left(\text{polylog}(n)\right)$	$O(\sqrt{\log n})$	Cai and Sauerwald (ICALP'17)

Table 1.1: Here K is the initial discrepancy. λ is the second largest eigenvalue of the round matrix. d is of order the maximum degree of G.

Our result looks weaker than previous ones because $O(\sqrt{\log n})$ looks worse. However, we should point out that the number of rounds is potentially much smaller especially when we have poorly connected graphs where $1-\lambda$ can be $O(1/n^2)$ [58]. Also in reality $O(\sqrt{\log n})$ is usually very close to a small contant.

With respect to **random walks** in dynamic graph models, one important work is done by Avin et al. [11], which studied the cover and mixing times in a dynamic graph model. They also showed an exponential lower bound for the hitting time, proving that we can never guarantee good results in arbitrary dynamic graph models. One recent work is done by Sauerwald and Zanetti [87], where they showed general mixing and hitting times for dynamic graph models, depending on the minimum probability in the stationary distribution. The common defect of these works is that they both assumed

that all the graphs in the dynamic graph model share the same stationary distributon for the random walks on them when dealing with mixing times. This assumption is good for their analyses but is relevantly unrealistic. In our work, we dropped this strong condition and derive various results in different regimes. We refer our readers to the Table 4.4 in Chapter 4 for more details. The following table compares our contribution with theirs.

Conditions	**Mixing Times**	**Reference**
Connected, d-regular Diameter h, Uniform π	$O(dhn \log n)$	Avin, Kouck and Lotker (RSA'18)
Connected, Uniform π	$O\left(n/\pi_\star\right)$	Sauerwald and Zanetti (ICALP'18)
Not necessarily connected No uniform π	$O(\log n)$ (dense graphs) No mixing or $\Omega(n)$ (sparse)	Cai, Sauerwald and Zanetti (SIROCCO'20)

Table 1.2: Here π is the stationary distribution of the random walk on the graph. $\pi_\star$ is the minimum probability in π.

Another similar line of work focuses on the random walks in random graphs [23, 32, 39, 51, 18]. Note that this can be confused with the walk in dynamic graph models. The main difference is that previous works chose one graph from the random graph model, then the random walk is applied on this fixed graph. Hence in these works, they mainly discuss the performance of the walk on the giant component in a random graph model, e.g., Erdős-Rényi model. In contrast, in our regime, the graph keeps changing. However, it is still worth studying the techniques in this topic because they provided a wide range of relevant toolkits.

1.5 Outline of thesis

We give a summary of the contents of each chapter:

- Chapter 2 contains necessary background knowledge for the thesis. Mathematical notations and definitions will be included. We need tools from probability, linear algebra, and spectral graph theory. More Markov chain theory background will also be presented there.

- Chapter 3 demonstrates the results of load balancing with random input tokens on each processor in the network. We show how fast the network can be balanced, i.e., how long it takes to make the difference between the maximum load and the

minimum load small enough.[1]

- Chapter 4 contains the *mixing times* results for random walks in our evolving graph model (edge-Markovian model). Intuitively, this shows how fast a random walk converges into a stationary state.[2]

- Chapter 5 contains some auxiliary results related to the mixing times, such as the *mixing times* results regarding the ℓ_∞ distance of the random walks in our evolving graph model.

- Chapter 6 is a conclusion of the thesis, providing a summary of the thesis, a discussion on the meanings of the work, and future possible directions.

In each chapter, we will start with a literature review of the topic, showing the importance of the work. Then the main mathematical proofs will be presented. Experimental results will be added to show that our theoretical results are consistent with what would happen in real scenarios. At the end of each chapter, we will give a discussion and a conclusion for the topic in that chapter.

[1] The load balancing results have been published in The 44th International Colloquium on Automata, Languages, and Programming (ICALP 2017).

[2] The random walk results have been published in The 27th International Colloquium on Structural Information and Communication Complexity (SIROCCO 2020).

Chapter 2

Background

In this chapter, we introduce some necessary mathematical definitions and theorems. Readers are assumed to have basic knowledge in probability, statistics, linear algebra and graph theory, such as expectations, variance, matrix, eigenvalues, nodes and edges, etc. We will also give necessary background knowledge on Markov chain theory because it plays an important role in this thesis.

2.1 Markov chains and random walks

In this section we present fundamental concepts which are necessary for this thesis in Markov chain theory. We will give the classical (that is, static) versions of the definitions. Later in our dynamic models, we have to tweak the corresponding concepts such as the mixing times of random walks.

A **Markov chain** is a process which moves among the elements of a set Ω in the following way: when the process is at a state $x \in \Omega$, the next position is chosen according to a fixed probability $\mathbf{P}(x, \cdot)$ depending only on x. More precisely, a sequence of random variables $(X_0, X_1, ...)$ is a Markov chain with the state space Ω and transition probability matrix $\mathbf{P}$ if for all $x, y \in \Omega$, for all t, $\mathbb{P}[X_{t+1} = y \mid X_t = x] = \mathbf{P}(x, y)$. At the beginning of the process, the distribution of the starting state is μ_0. As the process proceeds, the distribution of the state at time t is $\mu_t = \mu_0 \mathbf{P}^t$.

2.1.1 Random walks

We study discrete Markov chains where Ω is discrete and finite. A random walk process is just a concrete example of discrete Markov chains. Given a connected graph $G = (V, E)$, V is the state space of the random walk on the graph and E decides the transition matrix. The transition probability of a *simple* random walk on a graph from the node u to v is

defined as

$$\mathbf{P}(u,v) = \begin{cases} \frac{1}{\deg(u)} & u \neq v, \{u,v\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Intuitively the process jumps to one of the neighbours of u uniformly randomly and it will not move to non-neighbouring nodes.

For technical reasons, a variant of the standard random walks is usually used in the literature: *lazy* random walks, which means the process has $\frac{1}{2}$ probability to stay where it is. Hence the transition matrix is redefined as:

$$\mathbf{P}(u,v) = \begin{cases} \frac{1}{2} & u = v, \\ \frac{1}{2\deg(u)} & u \neq v, \{u,v\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

The benefit of this definition is that the matrix $\mathbf{P}$ is guaranteed to be positive semi-definite. It brings us non-negative eigenvalues of $\mathbf{P}$, which gives us some technical benefits.

In this thesis, the lazy version will be used in later chapters unless stated otherwise. One intuitive reason is that the simple version may not converge in some graphs. For example, the simple random walk cannot converge on bipartite graphs.

2.1.2 Mixing times of Markov chains

In Markov chain theory, there is an important concept: *mixing times*. Before we reveal its definition we need to do some preparations. We need to understand that the mixing time is the time taken to make the distribution of a Markov chain sufficiently close to its *stationary distribution*. Hence to define the mixing time we need: 1. a measure to check the difference between two probability distributions; 2. the definition of the stationary distribution.

Definition 2.1.1. *The total variation distance between two distributions μ and ν is*

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|.$$

Equivalently,

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{s \in \Omega} |\mu(s) - \nu(s)|.$$

Under some assumptions, after some time the probability distribution μ_t will remain nearly unchanged. The resulting distribution π is called the *stationary distribution* of this Markov chain. In other words, after some time T, we know that for all $t \geq T$, $\mu_t = \pi$.

Definition 2.1.2 (Stationary distribution)**.** *The stationary distribution π of a Markov chain on a state space Ω with the transition matrix $\mathbf{P}$ is a probability distribution that*

satisfies:

$$\pi = \pi \mathbf{P}.$$

Under the condition that the chain is irreducible, the stationary distribution is unique.

In this definition we mention a concept: irreducible Markov chains. A Markov chain $\mathbf{P}$ is called *irreducible* if for every $x, y \in \Omega$, there exists a t such that $\mathbf{P}^t(x, y) > 0$. It means that it is possible to get from any state to any other state using only transitions of positive probability.

Another important concept about Markov chains is: reversible Markov chains, which will be used later in the thesis. Essentially a Markov chain is called *reversible* if

$$\pi(x)\mathbf{P}(x, y) = \pi(y)\mathbf{P}(y, x),$$

for all $x, y \in \Omega$.

Now we can define the mixing time:

Definition 2.1.3 (Mixing times)**.** *Let* $\mathbf{P}$ *be the transition matrix of a Markov chain defined in the previous section, and let* π *denote its stationary distribution. Assuming the initial distribution of the random walk is* μ_0*, the mixing time is defined by:*

$$t_{\text{mix}}(\epsilon) := \min\{t : \left\|\mu_0 \mathbf{P}^t - \pi\right\|_{TV} \leq \epsilon\}. \tag{2.1}$$

Usually we require ϵ to be a very small value so that we can say the distance between the chain distribution and the stationary distribution is small enough .

Remark 2.1.4. *Note that the definitions in this section are modified to fit the purposes of this thesis. More rigorous and detailed concepts can be found in classical references in the literature, for example [58, 69].*

2.2 Concentration tools in probability theory

In probability theory, an important observation is that for independently distributed random variables, their sum usually concentrates around their mean with high probability. A well-known theorem is the law of large numbers. Below we give two similar but more general results.

Theorem 2.2.1 (Berry-Esseen [19, 37] for non-identical r.v.)**.** *Let* $X_1, X_2, ..., X_n$ *be independently distributed with* $\mathbb{E}[X_i] = 0$*,* $\mathbb{E}[X_i^2] = \text{Var}[X_i] = \sigma_i^2$*, and* $\mathbb{E}[|X_i|^3] = \rho_i < \infty$*. If* $F_n(x)$ *is the distribution of* $\frac{X_1 + ... + X_n}{\sqrt{\sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2}}$ *and* $\Phi(x)$ *is the standard normal distribution, then*

$$|F_n(x) - \Phi(x)| \leq C_0 \psi_0,$$

where $\psi_0 = \left(\sum_{i=1}^n \sigma_i^2\right)^{-3/2} \cdot \sum_{i=1}^n \rho_i$ *and* $C_0 > 0$ *is a constant.*

Theorem 2.2.2 (Hoeffding's Inequality [48]). *Consider a collection of independent random variables $X_i \in [a_i, b_i]$ with $i \in [n]$. Then for any number $\delta > 0$,*

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} X_i - \mathbb{E}\left[\sum_{i=1}^{n} X_i\right]\right| \geqslant \delta\right] \leqslant 2 \cdot \exp\left(\frac{-2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

2.3 Measure theory and martingale

Martingale is a very useful but complicated tool in probability theory. In this thesis we will use it to prove an important theorem (Theorem 3.5.3). To properly define it, we also need some basic concepts in measure theory.

2.3.1 Measure theory

Given a set Ω, a σ-algebra is a collection $\mathcal{F}$ of subsets satisfying

1. $\Omega \in \mathcal{F}$.
2. If $A_1, A_2, ...$ are elements of $\mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.
3. If $A \in \mathcal{F}$, then $A^c := \Omega \setminus A \in \mathcal{F}$.

Given a set Ω with a σ-algebra $\mathcal{F}$, a function $f : \Omega \to \mathbb{R}$ is called **measurable** if $f^{-1}(B)$ is an element of $\mathcal{F}$ for all open sets B. f is also called $\mathcal{F}$-measurable.

Given $(\Omega, \mathcal{F})$, a random variable X is a measurable function defined on Ω. A **filtration** $\{\mathcal{F}_t\}$ is a sequence of σ-algebras such that $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ for all t. We say $\{X_t\}$ is **adapted to** $\{\mathcal{F}_t\}$ if X_t is $\mathcal{F}_t$-measurable for all t.

2.3.2 Martingale

A **martingale** with respect to a filtration $\{\mathcal{F}_t\}$ is a sequence of random variables (M_t) satisfying the following conditions:

1. $\mathbb{E}[|M_t|] < \infty$ for all t.
2. (M_t) is adapted to $\{\mathcal{F}_t\}$.
3. $\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] = M_t$ for all $t \geq 0$.

Another important concept related to martingale is called a **stopping time**. A stopping time for the filtration $\{\mathcal{F}_t\}$ is a random variable τ with values in $\{0, 1, ...\}$ such that $\tau = t \in \mathcal{F}_t$.

2.4 Spectral graph theory

Spectral graph theory links the study of the graph properties and linear algebra. First we define a combinatorial measure of a given graph: *conductance*. It is a cut measure or a quantity that measures how well the graph can be divided into two clusters.

Definition 2.4.1. *The conductance of a non-empty set $S \subseteq V$ in a graph G is defined as:*

$$\Phi_G(S) := \frac{|E(S, V \setminus S)|}{\mathrm{vol}(S)},$$

where $\mathrm{vol}(S) := \sum_{x \in V} \deg(x)$. $\deg(x)$ *is the degree of the node x, which is the number of neighbours of x. $E(S, V \setminus S)$ is the set of edges between S and $V \setminus S$. The conductance of the entire graph G is defined as*

$$\Phi_G := \min_{\substack{S \subset V: \\ 1 \leq \mathrm{vol}(S) \leq \mathrm{vol}(V)/2}} \frac{|E(S, V \setminus S)|}{\mathrm{vol}(S)}.$$

The conductance of G and the second largest eigenvalue, λ_2, of the transition matrix P of a lazy random walk in G are related by the so-called discrete Cheeger inequality [7], which we state below.

Theorem 2.4.2 (Cheeger inequality)**.** *Let* $\mathbf{P}$ *be the transition matrix of a lazy random walk on a graph G. Then, it holds that*

$$1 - \lambda_2(\mathbf{P}) \leq \Phi_G \leq 2\sqrt{1 - \lambda_2(\mathbf{P})}.$$

Essentially this theorem connects the graph conductance, which is a graph property, with the second largest eigenvalue of the transition matrix of a lazy random walk on the same graph.

Note that the quantity $1 - \lambda_2$ is usually referred as the *spectral gap*. This is not just for the random walk. It can be also used for general Markov chains because it only depends on the transition matrix $\mathbf{P}$.

Chapter 3

Load Balancing with Random Inputs

3.1 Introduction of load balancing

In the last decade, large parallel networks have become widely available for industrial and academic users. An important prerequisite for their efficient usage is to balance their work efficiently. Load balancing is known to have applications to scheduling [92], routing [35], numerical computation such as solving partial differential equations [94, 93, 91], and finite element computations [50]. In the standard abstract formulation of load balancing, processors are represented by nodes of a graph, while links are represented by edges. The objective is to balance the load by allowing nodes to exchange loads with their neighbours via the incident edges.

Even if the total number of the tasks and the number of processors are known, we cannot easily arrange each processor to handle the average loads because the initial load distribution which can be very unbalanced. Passing a specific task to a far neighbour is time consuming and not worth doing. Due to these reasons, solving things locally is required, i.e., all processors only communicate with their neighbours simultaneously, to achieve a global balanced network. In this thesis, we will focus on *decentralized* and *iterative* load balancing protocols where a processor knows only its current load and the loads of the neighbouring processors. Based on that the processor decides how much load should be sent (or received).

In the rest of this chapter, the basic concepts of load balancing will be introduced first. Then we will present related works and summarise their contents to show the existing works of the literature. After that we demonstrate our main theorems and their proofs. In the end, the experiments are given to verify our theoretical results.

3.2 Background

3.2.1 Load balancing protocols

In load balancing, protocols are classified into different classes based on their assumptions on how processors transfer their loads. In this section, we introduce the concepts and delay the literature review to the Section 3.3.

1. **Diffusion model.** A widely used approach is diffusion, where load balancing algorithms distribute their loads to all their neighbours. In each step, the diffusive process happens in parallel on each node.

2. **Matching model.** We also consider the *matching model* (sometimes also called *dimension exchange model*), where in each round only the edges of a matching are used to average the load locally.

   In comparison to diffusion, the matching model reduces the communication in the network and moreover tends to behave in a more "monotone" fashion than diffusion, since it avoids concurrent load exchanges which may increase the maximum load or decrease the minimum load in certain cases. It contains two subclasses:

   - **Random matching model.** In each round a new random matching is generated by a simple distributed protocol, e.g., [43, 22].
   - **Balancing circuit model (a.k.a. dimension exchange).** A fixed sequence of matchings is applied periodically.

3.2.2 Metrics

To evaluate how good our load balancing protocols are, we apply two criteria: how small the *discrepancy* is and the number of *rounds* used to reach this discrepancy.

Discrepancy. We measure the smoothness of the load distribution by the so-called *discrepancy* which is the difference between the maximum and minimum load. In view of more complex scenarios where jobs are eventually removed or new jobs are generated, the discrepancy seems to be a more appropriate measure than the *makespan* which only considers the maximum load.

Rounds. We measure the time by the number of rounds used by a protocol to achieve our target discrepancy. In each round, some tokens are transferred among processors based on the chosen iterative protocol. The goal is to run the protocol as shortly as possible by design the protocol appropriately.

3.3 Related works

The origin of load balancing can trace back to [20, 42, 91, 3, 34]. Since then, a series of works have been studied, e.g., [75, 66, 13, 4, 17]. In this section, we first explore more in the diffusion model vs. the matching model. Then for both models, there are different types of cases with various assumptions. First we introduce the continuous case vs. the discrete case. Next in line is the number of tokens transferred among processors can be deterministic or randomised. Lastly, we demonstrate a relevantly new standard which has not appeared as much in the literature, i.e., the worst case vs. the average case. Another important line of work studies *selfish load balancing*, e.g., [14]. For the coherence of the thesis, we do not include them in this section for now. A review on that topic will be in Chapter 6.

Since each work can satisfy multiple assumptions, the following content may have duplicated citations. The main purpose is to convince readers into appreaciating the importance of these criteria. One reference may be only put in the section that shows its importance the most even if it can be put in multiple sections.

Diffusion model vs. Matching model. Early works first consider the diffusion models [3, 34]. Rabani et al. [75] created a solid foundation in recent load balancing research. They studied both the diffusion model and the balancing circuit model and get the same result for both. Hence this sheds light on the fact that the two models are essentially connected and in many cases they may have similar performaces.

Later an important improvement is [17] where the authors continued the work in [75] and analysed classes of deterministic algorithms that balance better than that in [75]. However they only did the diffusive part.

Regarding matching models, in [43] Ghosh and Muthukrishnan claimed that it is more efficient to send many tokens to one neighbour than the diffusive way. Also in practical, matching models are more feasible to implement because running diffusive protocols requires concurrency. If the network is large and complex, this would be extremely hard.

In general, the analysis techniques for diffusion and matching models are very different. For matching models, potential functions are used frequently, e.g., [43, 42]. A function is chosen to represent the state of the network, and then the problem is simplified to proving the potential decreases in each step. This, however, is very hard to be applied for diffusion models. Intuitively, this is because the diffusion can change the state of the network dramatically within one step. Diffusion models tend to be analysed by algebraic methods, e.g., [34, 91, 67].

Continuous case vs. Discrete case. In load balancing, whether the loads are arbitrarily divisible is an important standard to classify the problems. If it can be divided arbitrarily, it is called a *continuous case*. In that case, load balancing corresponds to a Markov chain on the graph and one can resort to a wide range of established techniques

to analyze the convergence speed [20, 42, 66]. In particular, the *spectral gap*, as described in the previous chapter, captures the time to reach a small discrepancy fairly accurately, e.g., see [88, 75] for the diffusion and see [22, 43] for the matching model.

However, in many applications the processor's load may consist of tasks which are not further divisible. That is why the continuous case has been also referred to as "idealized case" [75]. A natural way to model indivisible tasks is the *unit-size token model* where one assumes a smallest load entity, the unit-size token, and load is always represented by a multiple of this smallest entity. In the following, we will refer to the unit-size token model as the *discrete case*. [67] proved the first rigorous result for the discrete case in the diffusion model. A commonly used model is called Rotor-Router [17, 4]. In each round, every node distributes all of its tokens using the rotor. The rotor distributes the tokens until nothing remains. Usually protocols in the discrete case distribute tokens as evenly as possible. If it is not possible without splitting some tokens, the vertex redistributes the extra tokens among all its neighbours randomly, e.g., [13].

Researchers are keen on filling in the gaps between the continuous case and the discrete case. A number of works have been done in the field [5, 40]. This is very important in this field because usually the techniques used for these two models can be very different from each other. If we can connect them, then it means sometimes we can solve one case by tackling the other one and then bridging the gap when the original case is hard.

Deterministic case vs. Randomised case. The deterministic case usually refers to the protocol that assigns the token in a deterministic way. One classic deterministic load balancing work was done by Berenbrink et al. [17]. They analysed diffusive models in d-regular graphs, and showed that algorithms that are *cumulatively fair* can achieve a small discrepancy quickly.

Unlike the deterministic rounding in most previous works, [41, 85] analyzed randomised rounding based strategies, meaning that an excess token will be distributed uniformly at random among the two communicating nodes. The authors of [41] showed that such randomisation leads to a roughly square root of the achieved discrepancy in the same time. The authors of [85] proved that with this strategy the time to reach constant discrepancy in the discrete case is essentially the same as the corresponding time in the continuous case. Their results hold both for the random matching and the balancing circuit model.

Worst-case vs. Average-case. We study load balancing with the *average case* inputs in this thesis. As mentioned in the introduction chapter, we assume the loads are drawn from some probability distributions, hence we call it "average". Note that we do not call it the "random case" in order to distinguish it from the randomised case mentioned above. Another reason is that in our analysis the expectation of the probability distribution where the loads are drawn plays an important role. Hence we call it "average" to emphasise

that. The majority of previous works adopted the usual *worst-case* framework for deriving bounds on the load discrepancy [75]. That means any upper bound on the discrepancy holds for an arbitrary input, i.e., an arbitrary initial load vector. While it is of course very natural and desirable to have such general bounds, the downside is that for graphs with poor expansion like cycles or 2D-tori, the convergence is rather slow, i.e., quadratic or linear in the number of nodes n. This serves as a *motivation* to explore an average-case input. Specifically, we assume that the number of tokens at each node is sampled independently from a known distribution, e.g., uniform, binomial, geometric, Poisson and power law distributions. Our main results demonstrate that the convergence of the load vector is considerably quicker (measured by the time the protocol uses to achieve a small load discrepancy), especially on networks with slow convergence in the worst-case such as cycles and 2D-tori.

We point out that many related problems including scheduling on parallel machines or load balancing in a dynamic setting (meaning that jobs are continuously added and processed) have been studied under random inputs, e.g., [45, 9, 6]. To the best of our knowledge, only very few works have studied this question in iterative load balancing. One exception is [81], which investigated the performance of continuous load balancing on tori in the diffusion model. In contrast to this work, only upper bounds are given and they hold for the multiplicative ratio between maximum and minimum load rather than the discrepancy. Another related work [12] presented a distributed algorithm for community detection that is based on averaging a random $\{-1, 1\}$ initial load vector.

3.4 Preliminaries

In this section, we present more details for the previous definitions and their formal mathematical notations. We formalise our network as $G = (V, E)$, that is an undirected, connected graph with n nodes labelled in $[1, 2, ..., n]$. The maximum degree of G is denoted by $\Delta := \max_{u \in V} d(u)$. The notations $\mathbb{P}[\mathcal{E}]$ and $\mathbb{E}[X]$ denote the probability of an event $\mathcal{E}$ and the expectation of a random variable X, respectively. For any n-dimensional load vector x, $\mathrm{disc}(x) = \max_i x_i - \min_i x_i$ denotes the *discrepancy*. By $\frac{\mathbf{1}}{\mathbf{n}}$ we denote the vector with all values being $\frac{1}{n}$.

3.4.1 Load vector

For the continuous case where loads are divisible, let $\xi_0 \in \mathbb{R}^n$ be the initial load represented as a row vector, and in every round we apply our iterative load balancing protocol once. We let ξ_t denote the load vector after t rounds. For the discrete case, tokens are indivisible and unit-sized. Let $x_0 \in \mathbb{N}^n$ be the initial load vector and x_t denote the load vector after t rounds.

3.4.2 Load balancing protocols

All our load balancing protocols can be represented by a sequence of $n \times n$ matrices. In each round, we multiply the load vector ξ_{t-1} (or x_{t-1}) by a round matrix defined by the corresponding protocol (by conventions in the literature, we use $\mathbf{P}$ for the diffusion model and $\mathbf{M}$ for the matching model) to get ξ_t (or x_t).

1. **Diffusion model.** For any u and v, we define the *round matrix* $\mathbf{P}(u,v) = \frac{1}{\gamma\Delta}$ if $\{u,v\} \in E$, if $u = v$ then $\mathbf{P}(u,v) = 1 - \frac{\deg(u)}{\gamma\Delta}$. Here Δ is the maximum degree of the graph. $1/\gamma$ usually serves as a parameter to control the laziness of a random walk in Markov chain theory. Usually we choose $\gamma = 2$, which makes it correspond to the lazy version of the random walk introduced in the previous chapter.

   In our case we stick to regular graphs, so we may not need Δ here. We also define $\lambda(\mathbf{P})$ to be $\mathbf{P}$'s second largest eigenvalue (in absolute value, i.e., $|\lambda_2(\mathbf{P})|$).

2. **Matching model.** In the matching model, every two matched nodes in round t balance their load as evenly as possible. This can be expressed by a symmetric $n \times n$ matching matrix $\mathbf{M}_t$, where with slight abuse of notation we use the same symbol for the matching and the corresponding matching matrix. Formally, matrix $\mathbf{M}_t$ is defined by $\mathbf{M}_t(u,u) = 1/2$, $\mathbf{M}_t(v,v) = 1/2$ and $\mathbf{M}_t(u,v) = \mathbf{M}_t(v,u) = 1/2$ if $\{u,v\} \in \mathbf{M}_t \subseteq E$, and $\mathbf{M}_t(u,u) = \mathbf{M}_t(v,v) = 1$, $\mathbf{M}_t(u,v) = 0$ $(u \neq v)$ if u, v are not matched.

   In the *balancing circuit model*, its *round matrix* is made by a specific sequence of matchings applied periodically. More precisely, let $\mathbf{M}_1, \ldots, \mathbf{M}_d$ be a sequence of d matching matrices, also called *period* [1]. In our work, we assume that $d = O(1)$. Then in step $t \geq 1$, we apply the matching matrix $\mathbf{M}_t := \mathbf{M}_{((t-1) \mod d)+1}$. We define the *round matrix* by $\mathbf{M} := \prod_{s=1}^{d} \mathbf{M}_s$. If $\mathbf{M}$ is symmetric, we define $\lambda(\mathbf{M})$ to be its second largest eigenvalue (in absolute value). Following [75], if $\mathbf{M}$ is not symmetric (which is usually the case), we define $\lambda(\mathbf{M})$ as the second largest eigenvalue of the symmetric matrix $\mathbf{M} \cdot \mathbf{M}^T$, where $\mathbf{M}^T$ is the transpose of $\mathbf{M}$. We always assume that $\lambda(\mathbf{M}) < 1$, which is guaranteed to hold if the matrix $\mathbf{M}$ is irreducible. Since $\mathbf{M}$ is doubly stochastic, all powers of $\mathbf{M}$ are doubly stochastic. This eigenvalue plays an important role in our work. A natural choice for the d matching matrices is given by an edge coloring of G. There are various efficient distributed edge coloring algorithms, e.g. [71, 72].

   We demonstrate the balancing circuit model on specific topologies here. For *cycles*, we will consider the natural "Odd-Even" scheme meaning that for $\mathbf{M}_1$, the matching consists of all edges $\{j, (j+1) \pmod n\}$ for any odd j, while for $\mathbf{M}_2$, the matching

[1] Note that $d = O(1)$ may be different from the maximal degree (or degree) of the underlying graph.

has all edges $\{j, (j+1) \pmod n\}$ for any even j. More generally, for r-dimensional tori with vertex set $[0, n^{1/r} - 1]^r$, we will have $2 \cdot r$ matchings in total, meaning that for every dimension $1 \leq i \leq r$ we have two matchings along dimension i, similar to the definition of matchings for the cycle. For *hypercubes*, the canonical choice is dimension exchange consisting of $d = \log_2 n$ matching matrices $\mathbf{M}_i$ by $\mathbf{M}_i(u, v) = 1/2$ if and only if the bit representation of u and v differ only in bit i.

In the *random matching model*, one generates a random matching in each round. There are several different protocols to generate the matchings. Muthukrishnan and Ghosh [43] analyzed a protocol for d-regular graphs where in the first stage every node picks an incident edge independently with probability $\Theta(1/d)$. In the second stage, we consider the matching formed by all edges that are not incident to any other edge chosen in the first stage. A similar protocol also works for non-regular graphs [22]. These protocols have two properties for our analysis. First, for

$$p_{\min} := \min_{t \in \mathbb{N}} \min_{\{u,v\} \in E} \mathbb{P}\left[\{u, v\} \in \mathbf{M}_t\right],$$

we have $p_{\min} \geq c_{\min} \cdot \frac{1}{\Delta}$ for some constant $c_{\min}$. Second, random matchings generated in different rounds are mutually independent.

Remark 3.4.1. *Note that in this thesis, we would like to highlight that the reader should not confuse $\mathbf{P}^t$ with $\mathbf{P}_t$. When the time t shows as a superscript,it is the exponent in the matrix multiplication. When it is a subscript, it is the specific matrix at time t. In the published version of the load balancing results, $\mathbf{P}_t$ was written as $\mathbf{P}^{(t)}$ to be distinguished from $\mathbf{P}^t$. In this thesis we use $\mathbf{P}_t$ to make it consistent with the random walk and graph theory conventions so that we can reduce the abuse of some notations.*

3.4.3 Continuous case vs. Discrete case

In the continuous case, load is arbitrarily divisible. We simply do the multiplication, i.e., $\xi_t = \xi_{t-1}\mathbf{P}$ for the diffusion model and $\xi_t = \xi_{t-1}\mathbf{M}$ for the matching model.

When turning to the discrete case, let x_t be the load vector at the end of round t. For the diffusion model, the number of tokens transmitted is rounded down to the nearest integer [75]. For the matching model, when the sum of tokens of the two paired nodes is odd, we employ the *random orientation* (or *randomized rounding*) [75, 85]. More precisely, if there are two nodes u and v with load a and b being paired by matching $\mathbf{M}_t$, then node u gets either $\lceil \frac{a+b}{2} \rceil$ or $\lfloor \frac{a+b}{2} \rfloor$ tokens, with probability $1/2$ each. The remaining tokens are assigned to node v.

3.4.4 The average-case setting

We refer to average-case when each entry of the initial load vector is drawn from a fixed (discrete) distribution. In this thesis, we study the following distributions:

1. Exponentially concentrated distributions: uniform, binomial, geometric, Poisson distributions. These are commonly known distributions.

2. Non-exponentially concentrated distribution: power law distribution, which will be defined in Section 3.5.2.

Exponentially concentrated distributions are defined as below.

Definition 3.4.2. *A distribution D over $\mathbb{N} \cup \{0\}$ with expectation μ and variance σ^2 is* exponentially concentrated *if there is a constant $\kappa > 0$ so that for any $X \sim D$, $\delta > 0$,*

$$\mathbb{P}\left[|X - \mu| \geq \delta \cdot \sigma\right] \leq \exp\left(-\kappa\delta\right).$$

A Poisson distribution can be proved to be exponentially concentrated. Here we directly use a theorem from [8].

Theorem 3.4.3 ([8, Theorem A.1.15])**.** *Let X have a Poisson distribution with mean μ. Then for any $\epsilon > 0$,*

$$\mathbb{P}\left[X \leq (1-\epsilon)\mu\right] \leq e^{-\epsilon^2\mu/2},$$
$$\mathbb{P}\left[X \geq (1+\epsilon)\mu\right] \leq \left[e^{\epsilon}(1+\epsilon)^{-(1+\epsilon)}\right]^{\mu} \leq e^{-\min(\epsilon,\epsilon^2)\cdot\mu/3}.$$

Lemma 3.4.4. *The uniform distribution, binomial distribution, geometric distribution and Poisson distribution are all exponentially concentrated.*

Proof. Note that bounded distributions like the uniform distribution $\mathsf{Uni}[0,k]$ are trivially exponentially concentrated, since $\sigma = \Theta(k)$. The distributions with unbounded range may be exponentially concentrated, with one example being the geometric distribution $\mathsf{Geo}(p)$. To verify this, first note that we have $\mu = 1/p$ and $\sigma = \sqrt{(1-p)/p^2}$ (and so $\mu = \Theta(\sigma)$) and thus $\mathbb{P}\left[\mu - X \geq \delta \cdot \sigma\right] \leq \exp\left(-\kappa\delta\right)$ holds trivially for a sufficiently small constant $\kappa > 0$. Secondly, for the upper tail, by Markov's inequality, $\mathbb{P}\left[X \geq 2 \cdot \mathbb{E}\left[X\right]\right] \leq 1/2$, and by the memoryless property of the geometric distribution, for any integer $j \geq 1$, $\mathbb{P}\left[X \geq j \cdot 2 \cdot \mathbb{E}\left[X\right]\right] \leq 2^{-j}$.

For the binomial distribution $\mathsf{Bin}[m,p]$ with expectation $\mu = m \cdot p$ and standard deviation $\sigma = \sqrt{m \cdot p \cdot (1-p)}$, we will assume w.l.o.g. that $p \leq 1/2$, so that $\sigma = \Theta(\sqrt{mp})$. Then by [63, Theorem 2.3], we have for $X \sim \mathsf{Bin}[m,p]$, $\mathbb{P}\left[X - \mu \geq \epsilon \cdot \mu\right] \leq \exp\left(-\frac{\epsilon^2\mu}{2+2\epsilon/3}\right)$. Choosing $\epsilon = \delta \cdot \sigma/\mu$ yields $\mathbb{P}\left[X - \mu \geq \delta \cdot \sigma\right] \leq \exp\left(-\frac{\delta^2 \cdot \sigma^2/\mu}{2+2\sigma/\mu}\right)$, as needed.

For the lower tails, we use $\mathbb{P}[\mu - X \geq \epsilon \cdot \mu] \leq e^{-1/2\epsilon^2\mu}$ and obtain a similar result as before (see again [63, Theorem 2.3]).

For the Poisson Distribution $\mathsf{Poi}[\mu]$, we can verify in an analogous way that it is exponentially distributed by using two Chernoff bounds for Poisson random variables (Lemma 3.4.3). □

Therefore, our distributions can be partitioned into two simple classes: exponentially concentrated distributions without heavy tails and power law distributions with heavy tails (see also Definition 3.5.7).

Lemma 3.4.5. *Let D be an exponentially concentrated distribution and let $X \sim D$. Then, for a constant κ*

$$\mathbb{P}[|X - \mu| \leq 8/\kappa \cdot \sigma \log n] \geq 1 - n^{-3}.$$

In particular, the initial discrepancy satisfies $\operatorname{disc}(x_0) = O(\sigma \cdot \log n)$ *with probability* $1 - n^{-3}$.

The advantage of Lemma 3.4.5 is that we can use a simple conditioning trick to work with distributions that have a finite range and can therefore be analyzed by Hoeffding's inequality. Therefore in the analysis we may simply work with a bounded-range distribution $\widetilde{D}$, which is D under the condition that only values in the interval $[\mu - 8/\kappa \cdot \sigma \log n, \mu + 8/\kappa \cdot \sigma \log n]$ occur. Note that the constant 8 is not the smallest number to guarantee this probability, but it is sufficiently good and its influence is negligible (in asymptotic terms).

3.5 Main theorems

In this thesis, we study average case load balancing with discrete loads. Regarding the load balancing protocols, we analyze the diffusion model and the matching model together because they are more related. The underlying graph is assumed to have a bounded degree, i.e., the maximum degree has a fixed bound d. An example of unbounded graph models is hypercubes, where the degree is an unbounded function of n. As for the random matching model, the result can be proved similarly.

As mentioned before, we classify the types of the possible distributions into two classes: exponentially concentrated distributions and power law distribution. We will establish this classification in the next two subsections, respectively.

3.5.1 Exponentially concentrated distributions results

In this subsection, we present our theoretical results for the diffusion model and the matching model whose initial loads are drawn from independent and identically distributed ex-

ponentially concentrated distributions. The distribution has a mean of μ and a variance of σ^2.

The first result is a general formula that expresses the load difference between an arbitrary pair of nodes in round t for both the diffusion model and the balancing circuits.

Theorem 3.5.1. *Consider the diffusion model with an arbitrary round matrix* $\mathbf{P}$ *where the initial loads are drawn from an exponentially concentrated distribution. Then for any pair of nodes* u, v*, a constant* κ *as in Lemma (3.4.5) and round* t*, it holds for any* $\delta > 0$ *that*

$$\mathbb{P}\left[|x_t(u) - x_t(v)| \geq \frac{16\sqrt{2}\delta\sigma\log n}{\kappa} \cdot \left\|\mathbf{P}^t(.,u) - \mathbf{P}^t(.,v)\right\|_2 + \sqrt{64\log n}\right] \leq 2 \cdot e^{-\delta^2} + 2n^{-3}.$$

Furthermore, for any pair of vertices u, v*, we have the following lower bound:*

$$\mathbb{P}\left[|x_t(u) - x_t(v)| \geq \sigma/(2\sqrt{2\log_2 \sigma}) \cdot \left\|\mathbf{P}^t(.,u) - \mathbf{P}^t(.,v)\right\|_2 - \sqrt{64\log n}\right] \geq \frac{1}{16}.$$

Here $\mathbf{P}^t(.,u)$ is the u-th column vector in the matrix $\mathbf{P}^t$ which is the t-step transition matrix. Intuitively, $\mathbf{P}^t(w,u)$ is the portion of load that the node w transfers to the node u after t steps. So this theorem bounds the difference between the loads of any two nodes after round t. We delay the proof to Section 3.6.

Remark 3.5.2. *The lower bound above is not helpful if* σ *is small, say, at most a constant. However for sufficiently large* σ*, the lower bound gives a useful result (see also Section 3.10 & 3.11).*

For the matching model, the $\mathbf{P}$ in the above theorem should be replaced by the round matrix $\mathbf{M}$. For simplicity, we omit it here. One detail is that the time t for the balancing circuit model actually means that it applies the round matrix t times. However, in fact it is $t \cdot d$ matchings. Therefore, the time complexity for the balancing circuit model in this thesis actually differs from the real complexity by d. Since we have already assumed that d is a constant 3.4, this will not affect our results too much.

The proof of the upper bound in Theorem 3.5.1 relies on a previous result relating continuous and discrete load balancing from [85]. The lower bound is technically more challenging and applies a generalized version of the central limit theorem (Theorem 2.2.1).

Together, the upper and lower bound above establish that the load deviation between any two nodes u and v is essentially captured by the ℓ_2 norm $\|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2$. However, in some instances it might be desirable to have a more tangible estimate at the expense of generality. To this end, we first observe that $\|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2^2 \leq 4\max_{k\in V}\left\|\mathbf{M}^t(.,k) - \frac{\mathbf{1}}{\mathbf{n}}\right\|_2^2$ (see Lemma 3.7.1). Hence we are left with the problem of bounding the t-step probability vector $\mathbf{M}^t(.,k)$.

For reversible Markov chains, the upper bound has been analyzed in several works. For example, [61, Lemma 3.6] implies that for random walks on graphs, $\mathbf{P}^t(u,v) = O(\deg(v)/\sqrt{t})$. For the diffusion model, we can bound the t-step probability similarly.

However, the Markov chain associated to $\mathbf{M}$ for the matching model is not reversible in general. Bounding the t-step probability for the matching model is the most important part in this thesis. For the balancing circuits, we adopt the so-called evolving set process used in [59]. Specifically, they proved in [59, Theorem 17.17] that if $\mathbf{P}$ denotes the transition matrix of a lazy random walk (i.e., a random walk with loop probability at least $1/2$) on a graph with maximum degree Δ, π the stationary distribution of $\mathbf{P}$, then for any vertex $x \in V$:

$$\left|\mathbf{P}^t(x,x) - \pi(x)\right| \leq \frac{3\Delta^{5/2}}{\sqrt{t}}.$$

Such estimates have been used in applications besides load balancing, including distributed random walks and spanning tree enumeration [84, 61]. We generalize this result to Markov chains with an arbitrary loop probability and to arbitrary t-step transition probabilities. This is an important result gained on the way to our main result.

Theorem 3.5.3. *Let $\mathbf{P}$ be the transition matrix of an irreducible Markov chain and π its stationary distribution. Then we have for all states x, y and step t,*

$$\left|\mathbf{P}^t(x,y) - \pi(y)\right| \leq \frac{\pi_{\max}^{3/2}}{\pi_{\min}^{3/2}} \cdot \frac{2}{\beta^{1/2}\alpha}\sqrt{\frac{1-\beta+\alpha}{\alpha t}},$$

where $\alpha := \min_{u \neq v} \mathbf{P}(u,v) > 0$ and $\beta := \min_u \mathbf{P}(u,u) > 0$. $\pi_{\max}, \pi_{\min}$ are the maximum and minimum entries in the stationary distribution respectively.

Applying this bound to a round matrix $\mathbf{M}$ formed of $d = O(1)$ matchings for the balancing circuit model we obtain $|\mathbf{M}^t(u,v) - 1/n| = O(t^{-1/2})$. It should be noted that [85, Lemma 2.5] proved a weaker version where the upper bound is only $O(t^{-1/8})$ instead of $O(t^{-1/2})$. As discussed in Lemma 3.10.3, the bound $O(t^{-1/2})$ is asymptotically tight if we consider the balancing circuit model on poorly connected graphs like *cycles or tori*.

For the random matching model, we prove a similar $O(t^{-1/2})$ bound.

Theorem 3.5.4. *Consider the random matching model with a sequence of random matching matrices. Let $\mathbf{M}^t$ denote the product of the first t random matching matrices generated by the protocol. Then the expected value of $\|\mathbf{M}^t(u,.)\|_2^2$ is bounded by $O(t^{-1/2})$.*

Note that Theorem 3.5.3 works for the diffusion model and the balancing circuit model and Theorem 3.5.4 is used for the random matching model. We combine these t-step bounds with the upper bound in Theorem 3.5.1 to yield the following Corollary.

Corollary 3.5.5. *Consider the diffusion model and the matching model where the initial loads are drawn from an exponentially concentrated distribution. Then the load discrepancy after t rounds is $O(t^{-1/4} \cdot \sigma \cdot (\log n)^{3/2} + \sqrt{\log n})$ with probability $1 - O(n^{-1})$.*

Since the initial discrepancy is $O(\sigma \cdot \log n)$ (see Lemma 3.4.5), Corollary 3.5.5 implies that in the average case, there is a significant decrease (roughly of order $t^{-1/4}$) in the discrepancy, regardless of the underlying topology. For round matrices with small second largest eigenvalue, the next result provides a significant improvement:

Theorem 3.5.6. *Consider the diffusion model and the balancing circuit model with arbitrary round matrices* $\mathbf{P}, \mathbf{M}$ *with small second largest eigenvalues respectively where the initial loads are drawn from an exponentially concentrated distribution.* $\mathbf{M}$ *consists of* $d = O(1)$ *matchings. Let* λ *be* $\lambda_2(\mathbf{P})$ *for the diffusion model and* $\lambda_2(\mathbf{M})$ *for the balancing circuit model. Then the discrepancy after* t *rounds is* $O(\lambda^{t/4} \cdot \sigma \cdot (\log n)^{3/2} + \sqrt{\log n})$ *with probability* $1 - O(n^{-1})$.

Hence for graphs where λ is bounded away from 1, we even obtain an exponential convergence. In Section 3.10, we derive bounds on the discrepancy for concrete topologies (see Figure 3.1).

Note that we only mention the diffusion model and the balancing circuit model in this improved theorem. Essentially the balancing circuit model is more similar to the diffusion model than it to the random matching model because once the $n \times n$ matrices, $\mathbf{P}$ for the diffusion model and $\mathbf{M}$ for the balancing circuit model, are fixed, we just apply them to our load distribution every round. We *do not* include the random matching model because in each round the protocol generates a new random matching and we do not keep a fixed round matrix as we do when using the diffusion model and the balancing circuit model.

Graph	$\mathrm{disc}(x_t)$
Cycle	$t^{-1/4} \cdot \sigma$
r-dim. Torus	$t^{-r/4} \cdot \sigma$
Expander	$\lambda^{t/4} \cdot \sigma$
Hypercube	$2^{-t/2} \cdot \sigma$

Figure 3.1: Discrepancy bounds (without logarithmic factors) for different topologies. $\mathrm{disc}(x_t)$ is the discrepancy of x_t, which means the difference between the maximum and the minimum entries of x_t.

3.5.2 Non-exponentially concentrated distribution results (Power law distribution)

In contrast to exponentially concentrated distributions, we also study the power law distribution which has a heavy tail, meaning that its tail is not exponentially bounded. There are different types of power law distributions that share a general definition. However, to analyse it we need a more specific version of it. Here we give a formal definition of a power law distribution taken from [46] for this thesis:

Definition 3.5.7. *A power-law distribution, a.k.a. the zeta distribution or discrete Pareto distribution is defined as*

$$\mathbb{P}[X = x] = \frac{x^{-\alpha+1}}{\zeta(\alpha - 1)},$$

where $\alpha > 1$, x is a positive integer and ζ is the Riemann zeta function defined as $\sum_{k=1}^{\infty} k^{-\alpha+1}$.

In our work, we restrict $\alpha \in (2, 3)$ where the expectation exists but the variance is infinity. The case with the power law distribution is special because it has an unbounded range and its tail cannot be exponentially bounded. Hence concentration tools are not applicable. Instead of showing nearly matching lower and upper bounds on the discrepancy as in the exponentially concentrated cases, here we only present an upper bound on the expected discrepancy:

Theorem 3.5.8. *Consider the diffusion model and the balancing circuit model and the initial loads are drawn from i.i.d. power law distributions defined in Definition 3.5.7. The expected discrepancy is $O(\log(n)n^{1/(\alpha-1)}t^{-1/2})$.*

Later in Section 3.11, we discuss our results and compare them with the convergence of the discrepancy in the worst-case. On a high level, these results demonstrate that on all the considered topologies, we have much faster convergence in the average-case than in the worst-case. However, if we are only interested in the time to achieve a very small, say, constant or poly-logarithmic discrepancy, then we reveal an interesting dichotomy: we have a quicker convergence than in the worst-case iff the standard deviation σ (of the exponentially concentrated distribution) is smaller than some threshold depending on the topology. We observe the same phenomena in our experiments.

3.6 Proof of Theorem 3.5.1

Theorem 3.5.1 consists of an upper bound and a lower bound. We prove them separately.

3.6.1 Proof of Theorem 3.5.1 (Upper Bound)

We will use the following result from [85] that bounds the deviation between the continuous and discrete load, assuming that we have $\xi_0 = x_0$.

Theorem 3.6.1 ([85, Theorem 3.6(i)])**.** *For any sequence of matchings, let $x_0 = \xi_0$. Then for any round $t \geq 1$ it holds that*

$$\mathbb{P}\left[\max_{w \in V} |x_t(w) - \xi_t(w)| \leq \sqrt{16 \cdot \log n}\right] \geq 1 - 2n^{-3}.$$

The basic proof idea is as follows. Since $\xi_t(u) - \xi_t(v) = \sum_{w \in V} \xi_0(w) \cdot (\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v))$, it is a weighted sum of n i.i.d. random variables and its expectation is 0. We then can apply Hoeffding's inequality to obtain the theorem.

Proof of Theorem 3.5.1 (Upper Bound). Recall that the initial vector $\xi_0 = x_0$ consists of n i.i.d. random variables. As explained at the end of Section 3.4, we condition on the event

$$\mathcal{E} := \bigcap_{w \in V} \{|\xi_0(w) - \mu| \leq 8/\kappa \cdot \sigma \cdot \log n\}.$$

By Lemma 3.4.5, $\mathbb{P}[\mathcal{E}] \geq 1 - n^{-2}$. In the remainder of the proof, all random variables are conditional on $\mathcal{E}$, but for simplicity we will not explicitly express this conditioning.

Since $\xi_t(u) = \sum_{w \in V} \xi_0(w) \mathbf{M}^t(w,u)$, the load $\xi_t(u)$ is just a weighted sum of i.i.d. random variables and we obtain

$$\xi_t(u) - \xi_t(v) = \sum_{w \in V} \xi_0(w) \cdot \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right),$$

which is in fact still a weighted sum of n i.i.d. random variables. The expectation is

$$\begin{aligned}\mathbb{E}[\xi_t(u) - \xi_t(v)] &= \mathbb{E}\left[\sum_{w \in V} \xi_0(w) \cdot \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)\right] \\ &= \sum_{w \in V} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right) \mathbb{E}[\xi_0(w)] = 0,\end{aligned}$$

where the last equality holds since $\mathbf{M}$ is doubly stochastic and hence $\sum_{w \in V} \mathbf{M}^t(w,u) = \sum_{w \in V} \mathbf{M}^t(w,v) = 1$. Also $\mathbb{E}[\xi_0(w)]$ is the same for all $w \in V$ because they are all drawn from i.i.d. Thus we can take the expectation out and combine the terms.

By applying Hoeffding's inequality (Theorem 2.2.2) and recalling that conditional on $\mathcal{E}$, the range of each $\xi_t(w)$ is $16/\kappa \cdot \sigma \cdot \log n$, we obtain that

$$\mathbb{P}[|\xi_t(u) - \xi_t(v)| \geq \delta] \leq 2 \exp\left(\frac{-2\delta^2}{256/\kappa^2 \cdot \sigma^2 \log^2 n \, \|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2^2}\right).$$

Applying Theorem 3.6.1 yields

$$\begin{aligned}\mathbb{P}\left[|x_t(u) - x_t(v)| \geq \delta + \sqrt{64 \log n}\right] \leq 2 \exp&\left(\frac{-2\delta^2}{256/\kappa^2 \cdot \sigma^2 \log^2 n \, \|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2^2}\right) \\ &+ 4n^{-3}.\end{aligned}$$

The statement of the theorem follows by scaling δ and recalling that $\mathbb{P}[\mathcal{E}] \geq 1 - n^{-2}$. $\square$

3.6.2 Proof of Theorem 3.5.1 (Lower Bound)

The proof of the lower bound will use the following quantitative version of a central limit type theorem for independent but non-identical random variables.

With the concentration tool (Berry-Esseen) at hand, we are able to prove the lower bound in Theorem 3.5.1. Unfortunately, it appears quite difficult to apply Theorem 2.2.1 directly to $\xi_t(u)-\xi_t(v)$, since we need a good bound on the error term ψ_0. To this end, we will first partition the vertex set V into buckets with equal contribution to $\xi_t(u) - \xi_t(v)$. Then we will apply Theorem 2.2.1 to the bucket with the largest variance.

Proof of Theorem 3.5.1 (Lower Bound). We first consider the deviation $\xi_t(u) - \xi_t(v)$ for any u, v:

$$\text{dev} := \xi_t(u) - \xi_t(v) = \sum_{w \in V} \xi_0(w) \cdot \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right),$$

which is a weighted sum of i.i.d. random variables with expectation μ and variance σ^2. $\mathbb{E}[\text{dev}] = \sum_{w \in V} \mathbb{E}[\xi_0(w)] \cdot (\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)) = 0$ since $\mathbf{M}$ is a doubly stochastic matrix. We will first partition the above sum into buckets where the weights of the random variables are roughly the same in order to apply the Berry-Esseen theorem. More precisely, we will partition V into $\lceil 2\log_2 \sigma \rceil$ buckets, where for each i we have

$$V_i := \left\{w \in V : |\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)| \in (2^{-i-1}, 2^{-i}]\right\}, i \in [1, 2\log_2 \sigma - 1],$$

and

$$V_{\log_2 \sigma} := \left\{w \in V : \left|\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right| \leq \frac{1}{\sigma^2}\right\}.$$

Further, let us consider the variance of dev, σ^2_{dev}. Since $\text{Var}[aX] = a^2\text{Var}[X]$ and the inputs are independent random variables, we have

$$\sigma^2_{\text{dev}} = \sum_{w \in V} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2 \sigma^2.$$

Then by the pigeonhole principle there exists an index $1 \leq i \leq 2\log_2 \sigma$ such that

$$\sum_{w \in V_i} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2 \sigma^2 \geq \frac{1}{2\log_2 \sigma} \cdot \sigma^2_{\text{dev}}.$$

By eliminating σ^2 on both sides,

$$\sum_{w \in V_i} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2 \geq \frac{1}{2\log_2 \sigma} \cdot \sum_{w \in V} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2.$$

Case 1: If $i = 2\log_2 \sigma$, we can derive that for any u, v,

$$\begin{aligned}\left\|\mathbf{M}^t(V_i,u) - \mathbf{M}^t(V_i,v)\right\|_2^2 &\leq \left\|\mathbf{M}^t(V_i,u) - \mathbf{M}^t(V_i,v)\right\|_\infty \cdot \left\|\mathbf{M}^t(V_i,u) - \mathbf{M}^t(V_i,v)\right\|_1 \\ &\leq \frac{1}{\sigma^2} \cdot 2 = O(\sigma^{-2}).\end{aligned}$$

Then in Theorem 3.5.1 $\|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2$ is $O(\sqrt{\log_2 \sigma} \cdot \sigma^{-1})$ and the lower bound holds trivially becasue the event becomes a comparison between an absolute value and a negative value. The probability on the left hand side is 1.

Case 2: In the remainder of the proof we assume that $i \in [1, 2\log_2 \sigma]$. We now decompose dev into $dev = S + S^c$, where

$$S := \sum_{w \in V_i} \xi_0(w) \cdot \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right) \text{ and } S^c := \sum_{w \notin V_i} \xi_0(w) \cdot \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right).$$

We apply Theorem 2.2.1 to S. Before applying Theorem 2.2.1, we scale the original distribution to $\xi_0'(w) = \xi_0(w) - \mu$ to make the expectation be 0. In preparation for this, let us first upper bound ψ_0. Using the definition of exponentially concentrated distributions, it follows that for any constant k, the first k moments of $\xi_0'(w)$are all bounded from above by $O(\sigma^k)$. Hence,

$$\begin{aligned}\psi_0 &= \frac{\sum_{w \in V_i} \mathbb{E}\left[\left|\xi_0'(w) \cdot (\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v))\right|^3\right]}{\left(\sum_{w \in V_i} \mathbb{E}\left[\left(\xi_0'(w) \cdot (\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v))\right)^2\right]\right)^{3/2}} \\ &\leq \frac{O(1) \cdot \sum_{w \in V_i} \left|\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right|^3}{\left(\sum_{w \in V_i} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2\right)^{3/2}}.\end{aligned}$$

For any $w \in V_i$, $|\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)| \in (2^{-i-1}, 2^{-i}]$, we can simplify the above expression as follows:

$$\psi_0 = O\left(\frac{|V_i| \cdot 2^{-3i}}{|V_i|^{3/2} \cdot 2^{-3i}}\right) = O(|V_i|^{-1/2}).$$

Case 2.1: In the following, we will assume that $|V_i| \geq C_1$, where $C_1 > 0$ is a sufficiently large constant to be specified later. Since $\mathrm{Var}[aX] = a^2 \mathrm{Var}[X]$, we have

$$\begin{aligned}F_n(x) &= \mathbb{P}\left[\frac{\sum_{w \in V_i} \xi_0'(w) \cdot (\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v))}{\sigma\sqrt{\sum_{w \in V_i} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2}} \leq x\right] \\ &= \mathbb{P}\left[S - \mathbb{E}[S] \leq x\sigma\sqrt{\sum_{w \in V_i} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2}\right].\end{aligned}$$

We use C_0 which has appeared in Theorem 2.2.1. Since $|V_i| \geq C_1$, there is a constant

$C_2 \propto C_0/\sqrt{C_1} > 0$ such that $C_0 \cdot \psi_0 \leq C_2$ and

$$\mathbb{P}\left[S - \mathbb{E}[S] \geq x\sigma\sqrt{\sum_{w\in V_i}\left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2}\right] \geq \Phi(-x) - C_0\psi_0 \geq \Phi(-x) - C_2.$$

Now let $\Phi^c(x)$ denote the complement of the standard normal distribution. By using [1, Formula 7.1.13] and substitution we get:

$$\frac{1}{\sqrt{\pi}(x+\sqrt{x^2+2})e^{x^2}} < \Phi^c(x) \leqslant \frac{1}{\sqrt{\pi}(x+\sqrt{x^2+4/\pi})e^{x^2}}.$$

Hence by $\Phi(-x) = \Phi^c(x)$, choosing $x = 1$ and C_1 sufficiently large, the constant C_2 is therefore almost negligible in the lower bound formula. Then we have

$$\mathbb{P}\left[S - \mathbb{E}[S] \geq \sigma\sqrt{\sum_{w\in V_i}\left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2}\right] \geq \frac{1}{16}.$$

Similarly, we can derive that

$$\mathbb{P}\left[\mathbb{E}[S] - S \geq \sigma\sqrt{\sum_{w\in V_i}\left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2}\right] \geq \frac{1}{16}.$$

Hence, independent of what the value S^c is, there is still a probability of at least $1/16$ so that $|S + S^c| \geq \sigma/2 \cdot \sqrt{1/(2\log_2\sigma)} \cdot \sqrt{\sum_{w\in V}\left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2}$, which completes the proof for the case $|V_i| \geq C_1$. Intuitively, S has smaller variance and is more concentrated, and if we know the lower bound of the probability that S deviates from its expectation by its standard deviation, then we know that for S^c this probability can only be larger since it deviates more.

Case 2.2: The case $|V_i| < C_1$ is similar. The basic idea is not to apply Berry-Esseen but simply use the fact that any exponentially distributed random variable deviates from the expectation by $\Omega(\sigma)$ with constant probability.

Let X be an exponentially concentrated random variable. By definition, there exists a constant $\kappa > 0$ such that for any $\lambda > 0$, $\mathbb{P}[|X - \mu| \geq \lambda \cdot \sigma] \leq \exp(-\kappa \cdot \lambda)$. For simplicity let us assume that $\mathbb{E}[X] = \mu = 0$. Suppose for the sake of contradiction that $\mathbb{P}[|X| \geq \sigma/(100\tilde{\kappa}^3)] \leq \exp(-10/\kappa^2)$, where $\tilde{\kappa} := \max\{\kappa, 1\}$. Essentially, this assumes that with very low probability can the random variable deviate too far from the expectation. Here κ should make sure that the right hand side probability is very small. Then,

$$\mathbb{E}\left[X^2\right] = \sum_{i=1}^{\infty}\mathbb{P}\left[X^2 \geq i\right]$$

$$
\begin{aligned}
&\leq \sigma^2/(100\tilde{\kappa}^3) + 100\sigma^2/\kappa^2 \cdot \mathbb{P}\left[X^2 \geq \sigma^2/(100\tilde{\kappa}^3)\right] + \sum_{i=100\sigma^2/\kappa^2}^{\infty} \mathbb{P}\left[X^2 \geq i\right] \\
&\leq \sigma^2/100^2 + 100\sigma^2/\kappa^2 \cdot \mathbb{P}\left[|X| \geq \sigma/(100\tilde{\kappa}^3)\right] + \sum_{i=100\sigma^2/\kappa^2}^{\infty} \mathbb{P}\left[|X| \geq \lceil\sqrt{i}\rceil\right] \\
&\leq \sigma^2/100^2 + 100\sigma^2/\kappa^2 \cdot \exp(-10/\kappa^2) + \sum_{\lambda=100/\kappa^2}^{\infty} \sigma^2 \cdot \mathbb{P}\left[|X| \geq \sqrt{\lambda} \cdot \sigma\right] \\
&\leq \sigma^2/100^2 + \sigma^2/10 + \sigma^2 \cdot \sum_{\lambda=100/\kappa^2}^{\infty} \exp\left(-\kappa \cdot \sqrt{\lambda}\right) \\
&< \sigma^2,
\end{aligned}
$$

which is the desired contradiction since this should be equal. The second line shows a partition. The first term uses 1 to upper bound the probability for the first $\sigma^2/100\tilde{\kappa}^3$ terms in the summation. The second term uses $\sigma^2/(100\tilde{\kappa}^3)$ to upper bound the first $100\sigma^2/\kappa^2$.

To estimate the last term in the above derivation, we need the following derivation:

$$
\sum_{\lambda=1}^{\infty} \exp(-\kappa\sqrt{\lambda}) \leq \sum_{\lambda=1}^{\infty} \frac{24}{(\kappa\sqrt{\lambda})^4} \leq \frac{4\pi^2}{\kappa^4}.
$$

The first inequality is simply from the Taylor series which implies for any $x \geq 0$, $\exp(x) \geq x^4/24$. The second inequality relies on the fact that $\sum_{i=1}^{\infty} \frac{1}{i^2} = \pi^2/6$.

In conclusion, we have shown so far that for exponentially concentrated random variables the probability that it deviates from its mean is not exponentially bounded. Hence there exist constants $\lambda, c > 0$ such that

$$
\mathbb{P}\left[|X| \geq \lambda \cdot \sigma\right] \geq c.
$$

Suppose now w.l.o.g. we remove the absolute value,

$$
\mathbb{P}\left[X \geq \lambda \cdot \sigma\right] \geq c/2,
$$

the argument works in the same way if we have $\mathbb{P}\left[X \leq -\lambda \cdot \sigma\right] \geq c/2$. We then increase how far X deviates from its mean and the probability should decrease dramatically. Our claim is now that it must be also true that

$$
\mathbb{P}\left[X \leq -\lambda \cdot c/3 \cdot \sigma\right] \geq c',
$$

where $c' \ll c/2$ is another constant, which has to be chosen sufficiently small so that the following derivation works.

For the sake of *contradiction*, we assume the opposite, i.e.,

$$\mathbb{P}\left[X \leq -\lambda \cdot c/3 \cdot \sigma\right] < c'.$$

Then the expectation of X would be too large, which can be concluded as follows:

$$\begin{aligned}
\mathbb{E}\left[X\right] &= \sum_{k=-\infty}^{\infty} k \cdot \mathbb{P}\left[X = k\right] \\
&\geq \sum_{k=\lambda\cdot\sigma}^{\infty} k \cdot \mathbb{P}\left[X = k\right] + \sum_{k=-\lambda\cdot c/3\cdot\sigma}^{0} k \cdot \mathbb{P}\left[X = k\right] \\
&\quad + \sum_{k=-c''\cdot\sigma}^{-\lambda\cdot c/3\cdot\sigma} k \cdot \mathbb{P}\left[X = k\right] + \sum_{k=-\infty}^{-c''\cdot\sigma} k \cdot \mathbb{P}\left[X = k\right],
\end{aligned}$$

where $c'' > 0$ is a sufficiently large chosen constant so that the fourth term is sufficiently large (note that the fourth term is negative), say, it is at least $-\lambda c/64$. The constant $c' > 0$ must be chosen sufficiently small to compensate that (specifically, we may want $c' \leq \lambda c/(64c'')$). We continue to lower bound $\mathbb{E}\left[X\right]$:

$$\begin{aligned}
\mathbb{E}\left[X\right] &\geq \mathbb{P}\left[X \geq \lambda\sigma\right] \cdot \lambda\sigma - \lambda c/3 \cdot \sigma + \mathbb{P}\left[X \leq -\lambda c/3\sigma\right] \cdot (-c''\sigma) - \lambda c/64 \\
&\geq \lambda c/2 \cdot \sigma - \lambda c/3 \cdot \sigma - c'c''\sigma - \lambda c/64 \\
&> 0,
\end{aligned}$$

where the last inequality holds due to our choice of c'. However, $\mathbb{E}\left[X\right] > 0$ contradicts $\mathbb{E}\left[X\right] = 0$, which means that our assumption $\mathbb{P}\left[X \leq -\lambda \cdot c/3 \cdot \sigma\right] < c'$ was wrong and we must have $\mathbb{P}\left[X \leq -\lambda c/3 \cdot \sigma\right] \geq c'$. Hence overall there is a (sufficiently small constant) $c_5 > 0$ such that $\mathbb{P}\left[X \leq c_5 \cdot \sigma\right] \geq c_5$, and $\mathbb{P}\left[X \leq -c_5 \cdot \sigma\right] \geq c_5$. □

3.7 Proof of Corollary 3.5.5 (Universal upper bound for diffusion models and matching models)

In the previous section we proved that the deviation between the loads of two nodes u and v is essentially captured by $\|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2$. However, in some cases it might be hard to compute or estimate this quantity for arbitrary vertices u and v. Therefore we will establish Corollary 3.5.5 which gives a more concrete estimate.

The proof of Corollary 3.5.5 is fairly involved and we sketch the high level ideas before giving the formal proof. We first show that $\|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2^2$ can be upper bounded in terms of the ℓ_2-distance to the stationary distribution. Then we will combine that with a direct application of a general bound on the t-step probabilities of an arbitrary,

possibly non-reversible Markov chain, as given in Theorem 3.5.3 from page 27 whose proof is deferred to Section 3.7.1.

Lemma 3.7.1. *Consider the balancing circuit model with an arbitrary round matrix* $\mathbf{M}$. *Then for all* $u, v \in V$, *we have* $\|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2^2 \leq 4 \cdot \max_{k \in V} \|\mathbf{M}^t(.,k) - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2$. *Further, for any* $u \in V$ *we have* $\max_{v \in V} \|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2^2 \geq \|\mathbf{M}^t(.,u) - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2$.

Proof. We expand the ℓ_2 norm based on its definition.

$$\begin{aligned}\sum_{w \in V} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2 &\leq 2 \cdot \left(\sum_{w \in V} \left(\mathbf{M}^t(w,u) - \frac{1}{n}\right)^2 + \left(\mathbf{M}^t(w,v) - \frac{1}{n}\right)^2\right) \\ &\leq 4 \cdot \max_{k \in V} \sum_{w \in V} \left(\mathbf{M}_t(w,k) - \frac{1}{n}\right)^2,\end{aligned}$$

and the first statement follows. We now prove the second statement:

$$\forall u \colon \max_{v \in V} \sqrt{\sum_{w \in V} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2} \geq \sqrt{\sum_{w \in V} \left(\mathbf{M}^t(w,u) - \frac{1}{n}\right)^2}.$$

We first look at the difference between these two terms squared. That is, for any vertex $v \in V$ we have

$$\begin{aligned}&\sum_{w \in V} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)^2 - \sum_{w \in V} \left(\mathbf{M}^t(w,u) - \frac{1}{n}\right)^2 \\ &= -2 \sum_{w \in V} \left(\mathbf{M}^t(w,u) - \frac{1}{n}\right)\left(\mathbf{M}^t(w,v) - \frac{1}{n}\right) + \sum_{w \in V} \left(\mathbf{M}^t(w,v) - \frac{1}{n}\right)^2 \\ &= -2 \sum_{w \in V} \mathbf{M}^t(w,u) \cdot \mathbf{M}^t(w,v) + \frac{4}{n} - \frac{2}{n} + \sum_{w \in V} \left(\mathbf{M}^t(w,v)\right)^2 + \frac{1}{n} \\ &= -2 \sum_{w \in V} \mathbf{M}^t(w,u) \cdot \mathbf{M}^t(w,v) + \sum_{w \in V} \left(\mathbf{M}^t(w,v)\right)^2 + \frac{1}{n}. \qquad (3.1)\end{aligned}$$

Now let Z be a uniform random variable over the set $V \backslash \{u\}$. Then it follows that

$$\begin{aligned}\mathbb{E}_{Z \sim V \backslash \{u\}} \left[\sum_{w \in V} \mathbf{M}^t(w,u) \cdot \mathbf{M}^t(w,Z)\right] &= \sum_{z \in V, z \neq u} \frac{1}{n-1} \cdot \sum_{w \in V} \mathbf{M}^t(w,u) \cdot \mathbf{M}^t(w,Z) \\ &= \frac{1}{n-1} \sum_{w \in V} \mathbf{M}^t(w,u) \cdot \sum_{z \in V, z \neq u} \mathbf{M}^t(w,Z) \\ &= \frac{1}{n-1} \sum_{w \in V} \mathbf{M}^t(w,u) \cdot \left(1 - \mathbf{M}^t(w,u)\right) \\ &= \frac{1}{n-1} \left(1 - \sum_{w \in V} \left(\mathbf{M}^t(w,u)\right)^2\right).\end{aligned}$$

Further, by linearity of expectations

$$\begin{aligned}
&\mathbb{E}_{Z\sim V\setminus\{u\}}\left[-2\sum_{w\in V}\mathbf{M}^t(w,u)\cdot\mathbf{M}^t(w,Z)+\sum_{w\in V}\left(\mathbf{M}^t(w,Z)\right)^2+\frac{1}{n}\right]\\
&=\frac{-2}{n-1}\left(1-\sum_{w\in V}\left(\mathbf{M}^t(w,u)\right)^2\right)+\sum_{z\in V,y\neq u}\sum_{w\in V}\frac{1}{n-1}\left(\mathbf{M}^t(w,Z)\right)^2+\frac{1}{n}.
\end{aligned}$$

By definition of expectation, this implies that there exists a vertex $v\in V, v\neq u$ such that

$$\begin{aligned}
&-2\sum_{w\in V}\mathbf{M}^t(w,u)\cdot\mathbf{M}^t(w,v)+\sum_{w\in V}\left(\mathbf{M}^t(w,v)\right)^2+\frac{1}{n}\\
&\geq\frac{-2}{n-1}\left(1-\sum_{w\in V}\left(\mathbf{M}^t(w,u)\right)^2\right)+\sum_{z\in V,z\neq u}\sum_{w\in V}\frac{1}{n-1}\left(\mathbf{M}^t(w,Z)\right)^2+\frac{1}{n}. \qquad (3.2)
\end{aligned}$$

Combining (3.1) and (3.2),

$$\begin{aligned}
&\sum_{w\in V}\left(\mathbf{M}^t(w,u)-\mathbf{M}^t(w,v)\right)^2-\sum_{w\in V}\left(\mathbf{M}^t(w,v)-\frac{1}{n}\right)^2\\
&\geq\frac{-2}{n-1}\cdot\left(1-\sum_{w\in V}\left(\mathbf{M}^t(w,u)\right)^2\right)+\sum_{w\in V}\sum_{z\in V,z\neq u}\frac{1}{n-1}\cdot\left(\mathbf{M}^t(w,Z)\right)^2+\frac{1}{n}\\
&=-\frac{1}{n-1}-\frac{1}{n\cdot(n-1)}+\frac{2}{n-1}\sum_{w\in V}\left(\mathbf{M}^t(w,u)\right)^2+\frac{1}{n-1}\cdot\sum_{w\in V}\sum_{z\in V,z\neq u}\left(\mathbf{M}^t(w,Z)\right)^2\\
&=-\frac{1}{n-1}-\frac{1}{n\cdot(n-1)}+\frac{1}{n-1}\sum_{w\in V}\left(\mathbf{M}^t(w,u)\right)^2+\frac{1}{n-1}\cdot\sum_{w\in V}\sum_{z\in V}\left(\mathbf{M}^t(w,Z)\right)^2\\
&\geq-\frac{1}{n-1}-\frac{1}{n\cdot(n-1)}+\frac{1}{n\cdot(n-1)}+\frac{1}{n-1}\\
&\geq 0,
\end{aligned}$$

where the second last inequality holds since $\mathbf{M}$ is doubly stochastic. □

The next step and main ingredient of the proof of Corollary 3.5.5 is to establish that $\|\mathbf{M}^t(.,k)-\frac{\mathbf{1}}{\mathbf{n}}\|_\infty= O(1/\sqrt{t})$. This result will be a direct application of Theorem 3.5.3.

Proof of Corollary 3.5.5. By Theorem 3.5.1 and Lemma 3.7.1, we obtain

$$\mathbb{P}\left[|x_t(u)-x_t(v)|\geq\frac{16\sqrt{2}\delta\sigma\log n}{\kappa}\max_{k\in V}\left\|\mathbf{M}^t(.,k)-\frac{\mathbf{1}}{\mathbf{n}}\right\|_2+\sqrt{64\log n}\right]\leq 2e^{-\delta^2}+2n^{-3}.$$

where $\mathbf{M}_t$ can either be the matrix for the balancing circuit model or the diffusion model.

Hence we can find a $\delta=\sqrt{3\log n}$ so that the latter probability gets smaller than $4n^{-3}$. Further, by applying Theorem 3.5.3 with $\alpha=\beta=2^{-d}$ to $\mathbf{P}=\mathbf{M}$ we conclude that for

the balancing circuit model $\|\mathbf{M}^t(.,k) - \frac{\mathbf{1}}{\mathbf{n}}\|_\infty = O(t^{-1/2})$, since $d = O(1)$. For the random matching model, we just insert Theorem 3.5.4. Using $\|.\|_2^2 \leq \|.\|_\infty \cdot \|.\|_1$, $\|\mathbf{M}^t(.,k) - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2 = O(t^{-1/2})$ and the union bound, $\mathrm{disc}(x_t) = O(t^{-1/4} \cdot \sigma \cdot (\log n)^{3/2} + \sqrt{\log n})$ with probability at least $1 - 4n^{-1}$. □

3.7.1 Proof of Theorem 3.5.3

As mentioned in the previous proof, Theorem 3.5.3 is very important to upper bound the t-step transition probability to $O(1/\sqrt{t})$. This section is devoted to completing the missing proof of Theorem 3.5.3.

Our proof is based on the evolving-set process, which is a Markov chain based on any given irreducible, not necessarily reversible Markov chain on Ω. For the definition of the evolving set process, we closely follow the exposition in [59, Chapter 17].

Let $\mathbf{P}$ denote the transition matrix of an irreducible Markov chain and π its stationary distribution. $\mathbf{P}^t$ is the t-step transition probability matrix. The *edge measure* Q is defined by $Q_{x,y} := \pi(x)\mathbf{P}(x,y)$ and $Q(A,B) = \sum_{x \in A, y \in B} Q_{x,y}$.

Definition 3.7.2. *Given a transition matrix $\mathbf{P}$, the **evolving-set process** is a Markov chain on subsets of Ω defined as follows. Suppose the current state is $S \subset \Omega$. Let U be a random variable which is uniform on $[0,1]$. The next state of the chain is the set*

$$\tilde{S} = \left\{ y \in \Omega : \frac{Q(S,y)}{\pi(y)} \geq U \right\}.$$

This chain is *not irreducible* because $\varnothing$ and Ω are absorbing states. It follows that

$$\mathbb{P}\left[y \in S_{t+1} \,|\, S_t\right] = \frac{Q(S_t,y)}{\pi(y)},$$

since the probability that $y \in S_{t+1}$ is equal to the probability of the event that the chosen value of U is less than $\frac{Q(S_t,y)}{\pi(y)}$.

Proposition 3.7.3 ([59, Proposition 17.19])**.** *Let (M_t) be a non-negative martingale with respect to (Y_t), and define $T_h := \min\{t \geq 0 : M_t = 0 \text{ or } M_t \geq h\}$ Assume that for any $h \geq 0$*

1. *For $t < T_h$, $\mathrm{Var}\left[M_{t+1} \,|\, Y_0, \ldots, Y_t\right] \geq \sigma^2$, and*
2. *$M_{T_h} \leq Dh$.*

Let $T := T_1$. If M_0 is a constant, then $\mathbb{P}\left[T > t\right] \leq \frac{2M_0}{\sigma}\sqrt{\frac{D}{t}}$.

The definition of the martingale can be found in Section 2.3. We now generalize [59, Lemma 17.14] to cover arbitrarily small loop probabilities.

Lemma 3.7.4. *Let (U_t) be a sequence of independent random variables, each uniform on $[0,1]$, such that S_{t+1} is generated from S_t using U_{t+1}. Then with $\beta := \min_u \mathbf{P}(u,u) > 0$,*

$$\mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} \le \beta, S_t = S\right] \ge \pi(S) + Q(S, S^c),$$
$$\mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} > \beta, S_t = S\right] \le \pi(S) - \frac{\beta Q(S, S^c)}{1-\beta}.$$

We list a few auxiliary results from [59] about the evolving set process that will be used to prove this lemma.

Lemma 3.7.5 ([59, Lemma 17.12])**.** *If $(S_t)_{t\ge 0}$ is the evolving-set process associated to the transition matrix $\mathbf{P}$, then for any time t and $x, y \in \Omega$*

$$\mathbf{P}^t(x,y) = \frac{\pi(y)}{\pi(x)} \mathbb{P}_{\{x\}}\left[y \in S_t\right].$$

Recall that (S_t) is the evolving-set process based on the Markov chain whose transition matrix is $\mathbf{P}$. $\mathbb{P}_{\{x\}}[y \in S_t]$ means the probability of the event $y \in S_t$ with the initial state of the evolving set being $\{x\}$.

Lemma 3.7.6 ([59, Lemma 17.13])**.** *The sequence $\{\pi(S_t)\}$ is a martingale.*

Theorem 3.7.7 ([59, Corollary 17.7])**.** *Let (M_t) be a martingale and τ a stopping time. If $\mathbb{P}[\tau < \infty]$ and $|M_{t\wedge\tau}| \le K$ for all t and some constant K where $t \wedge \tau := \min\{t, \tau\}$, then $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$.*

Proof of Lemma 3.7.4. Given $U_{t+1} \le \beta$, the distribution of U_{t+1} is uniform on $[0, \beta]$.
Case 1: For $y \notin S$, we know that for y satisfying $\frac{Q(S,y)}{\pi(y)} \in [0, \beta]$

$$\mathbb{P}\left[\frac{Q(S,y)}{\pi(y)} \ge U_{t+1} \,\middle|\, U_{t+1} \le \beta, S_t = S\right] = \frac{Q(S,y)}{\beta\pi(y)},$$

and for y satisfying $\frac{Q(S,y)}{\pi(y)} \in (\beta, 1]$,

$$\mathbb{P}\left[\frac{Q(S,y)}{\pi(y)} \ge U_{t+1} \,\middle|\, U_{t+1} \le \beta, S_t = S\right] = 1.$$

We know that

$$\frac{Q(S,y)}{\pi(y)} = \frac{\sum_{x\in S} \pi(x)\mathbf{P}(x,y)}{\pi(y)} \le \frac{\sum_{x\in\Omega} \pi(x)\mathbf{P}(x,y)}{\pi(y)} = 1.$$

Since $y \in S_{t+1}$ if and only if $U_{t+1} \le Q(S_t, y)/\pi(x)$, we therefore can combine the above results by using an inequality and conclude that

$$\mathbb{P}\left[y \in S_{t+1} \,|\, U_{t+1} \le \beta, S_t = S\right] \ge \frac{Q(S,y)}{\pi(y)} \text{ for } y \notin S.$$

because $\beta \leq 1$ and $Q(S,y)/\pi(y) \leq 1$.

Case 2: For $y \in S$, we have $Q(S,y)/\pi(y) \geq Q(y,y)/\pi(y) \geq \beta$, it follows that when $U_{t+1} \leq \beta$

$$\mathbb{P}\left[y \in S_{t+1} \,|\, U_{t+1} \leq \beta, S_t = S\right] = 1 \text{ for } y \in S.$$

We have

$$\begin{aligned}
&\mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} \leq \beta, S_t = S\right] \\
&= \mathbb{E}\left[\sum_{y\in\Omega} \mathbb{1}_{\{y\in S_{t+1}\}}\pi(y) \,\middle|\, U_{t+1} \leq \beta, S_t = S\right] \\
&= \sum_{y\in S} \pi(y)\mathbb{P}\left[y \in S_{t+1} \,|\, U_{t+1} \leq \beta, S_t = S\right] + \sum_{y\notin S} \pi(y)\mathbb{P}\left[y \in S_{t+1} \,|\, U_{t+1} \leq \beta, S_t = S\right].
\end{aligned}$$

Based on previous results, we can see that

$$\mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} \leq \beta, S_t = S\right] \geq \pi(S) + Q(S, S^c).$$

By Lemma 3.7.6 and the formulas above,

$$\begin{aligned}
\pi(S) &= \mathbb{E}\left[\pi(S_{t+1}) \,|\, S_t = S\right] \\
&= \beta \cdot \mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} \leq \beta, S_t = S\right] + (1-\beta) \cdot \mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} > \beta, S_t = S\right].
\end{aligned}$$

Rearranging shows that

$$\mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} > \beta, S_t = S\right] \leq \pi(S) - \frac{\beta Q(S, S^c)}{1-\beta}.$$

□

The derivation of the next lemma closely follows the analysis in [59, Chapter 17].

Theorem 3.7.8 (Optional Stopping Theorem [59, Corollary 17.7])**.** *Let (M_t) be a martingale and τ a stopping time. If $\mathbb{P}[\tau < \infty]$ and $|M_{t\wedge\tau}| \leq K$ for all t and some constant K where $t \wedge \tau := \min\{t, \tau\}$, then $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$.*

Lemma 3.7.9. *For any two states x, y, $|\mathbf{P}^t(x,y) - \pi(y)| \leq \frac{\pi(y)}{\pi(x)} \cdot \mathbb{P}_{\{x\}}[\tau > t]$.*

Proof. First of all, let the hitting time

$$\tau = \min\{t \geq 0 : S_t \in \{\varnothing, \Omega\}\}.$$

We have $S_\tau \in \{\varnothing, \Omega\}$ and $\pi(S_\tau) = \mathbb{1}_{\{S_\tau = \Omega\}}$. We consider an evolving set process with

$S_0 = \{x\}$. By Theorem 3.7.8 and Lemma 3.7.6,

$$\pi(x) = \mathbb{E}_{\{x\}}[\pi(S_0)] = \mathbb{E}_{\{x\}}[\pi(S_\tau)] = \mathbb{E}_{\{x\}}\left[\mathbb{1}_{\{S_\tau=\Omega\}}\right] = \mathbb{P}_{\{x\}}[S_\tau = \Omega] = \mathbb{P}_{\{x\}}[x \in S_\tau]. \tag{3.3}$$

For the last equality, it is true because we know that S_τ can only be $\varnothing$ or Ω. Hence, the probability that x is an element in S_τ is equal to the probability that S_τ is Ω. Note that here the second x in the last line can be any other element in Ω. For example, we also know that

$$\forall y \in \Omega, \mathbb{P}_{\{x\}}[S_\tau = \Omega] = \mathbb{P}_{\{x\}}[y \in S_\tau]. \tag{3.4}$$

For our bound, we know that by Lemma 3.7.5 and (3.3),

$$\begin{aligned}\left|\mathbf{P}^t(x,y) - \pi(y)\right| &= \frac{\pi(y)}{\pi(x)}\left|\mathbb{P}_{\{x\}}[y \in S_t] - \pi(x)\right| \\ &= \frac{\pi(y)}{\pi(x)}\left|\mathbb{P}_{\{x\}}[y \in S_t] - \mathbb{P}_{\{x\}}[S_\tau = \Omega]\right|.\end{aligned}$$

By (3.4),

$$\begin{aligned}\mathbb{P}_{\{x\}}[y \in S_t] &= \mathbb{P}_{\{x\}}[y \in S_t, \tau > t] + \mathbb{P}_{\{x\}}[y \in S_t, \tau \le t] \\ &= \mathbb{P}_{\{x\}}[y \in S_t, \tau > t] + \mathbb{P}_{\{x\}}[S_\tau = \Omega, \tau \le t].\end{aligned} \tag{3.5}$$

By simple substitution we obtain

$$\begin{aligned}\left|\mathbf{P}^t(x,y) - \pi(y)\right| &= \frac{\pi(y)}{\pi(x)}\left|\mathbb{P}_{\{x\}}[y \in S_t, \tau > t] + \mathbb{P}_{\{x\}}[S_\tau = \Omega, \tau \le t] - \mathbb{P}_{\{x\}}[S_\tau = \Omega]\right| \\ &= \frac{\pi(y)}{\pi(x)}\left|\mathbb{P}_{\{x\}}[y \in S_t, \tau > t] - \mathbb{P}_{\{x\}}[S_\tau = \Omega, \tau > t]\right| \\ &\le \frac{\pi(y)}{\pi(x)}\mathbb{P}_{\{x\}}[\tau > t].\end{aligned}$$

The last line is true because we remove all possible intersections. □

Now we want to use Proposition 3.7.3 to bound $\mathbb{P}_{\{x\}}[\tau > t]$. To apply it, we substitute the following parameters: M_0 is chosen to be $\pi(\{x\})$, Y_t is S_t, and $T = T_1 := \min\{t \ge 0 : \pi(S_t) = 0 \text{ or } \pi(S_t) \ge 1\}$. Hence in our case, τ is the same as T (or T_1) in the proposition. The following two lemmas elaborate on the two preconditions (i) and (ii) of Proposition 3.7.3.

Lemma 3.7.10. *For any time t and $S_0 = \{x\}$, $\mathrm{Var}_{S_t}[\pi(S_{t+1})] \ge \beta\pi_{\min}^2\alpha^2$.*

Proof. Conditioning always reduces variance and $S_t \neq \varnothing$ or Ω, we have

$$\mathrm{Var}_{S_t}[\pi(S_{t+1})] \ge \mathrm{Var}_{S_t}\left[\mathbb{E}\left[\pi(S_{t+1}) \mid \mathbb{1}_{\{U_{t+1}\le\beta\}}\right]\right].$$

For $S_t = S$,

$$\mathbb{E}_{S_t}\left[\pi(S_{t+1}) \,|\, \mathbb{1}_{\{U_{t+1}\leq\beta\}}\right] = \begin{cases} \mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} \leq \beta, S_t = S\right], & \text{w.p. } \beta, \\ \mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} > \beta, S_t = S\right], & \text{w.p. } 1-\beta, \end{cases}$$

and by Lemma 3.7.4, we know that

$$\mathbb{E}_{S_t}\left[\pi(S_{t+1}) \,|\, \mathbb{1}_{\{U_{t+1}\leq\beta\}}\right] \begin{cases} \geq \pi(S) + Q(S, S^c), & \text{w.p. } \beta, \\ \leq \pi(S) - \frac{\beta Q(S,S^c)}{1-\beta}, & \text{w.p. } 1-\beta. \end{cases}$$

For simplicity, we let $\mathbb{E}_{S_t}\left[\pi(S_{t+1}) \,|\, \mathbb{1}_{\{U_t\leq\beta\}}\right]$ be X, $\mathbb{E}\left[\pi(S_{t+1}) \,|\, U_{t+1} \leq \beta, S_t = S\right]$ be x_1 and $\mathbb{E}\left[\pi(S_{t+1} \,|\, U_{t+1} > \beta, S_t = S)\right]$ be x_2. Then we have

$$\begin{aligned} \mathrm{Var}_{S_t}\left[\mathbb{E}\left[\pi(S_{t+1}) \,|\, \mathbb{1}_{\{U_t\leq\beta\}}\right]\right] &= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2 \\ &= \beta x_1^2 + (1-\beta)x_2^2 - (\beta x_1 + (1-\beta)x_2)^2 \\ &= (\beta - \beta^2)(x_1 - x_2)^2. \end{aligned}$$

In order to derive a lower bounds on this variance, based on Lemma 3.7.4 we let $x_1 = \pi(S) + Q(S, S^c)$ and $x_2 = \pi(S) - (\beta/1-\beta)Q(S, S^c)$. With this we obtain

$$\mathrm{Var}_{S_t}\left[\mathbb{E}\left[\pi(S_{t+1}) \,|\, \mathbb{1}_{\{U_t\leq\beta\}}\right]\right] \geq \frac{\beta}{1-\beta} Q^2(S, S^c).$$

Therefore, provided $S_t \notin \{\varnothing, \Omega\}$, we have

$$\mathrm{Var}_{S_t}\left[\pi(S_{t+1})\right] \geq \frac{\beta}{1-\beta} Q^2(S, S^c) \geq \frac{\beta \pi_{\min}^2 \alpha^2}{(1-\beta)}.$$

The last inequality follows from the fact that if $S \notin \{\varnothing, \Omega\}$ then there exist $u \in S, v \notin S$ with $\mathbf{P}(u, v) > 0$, whence

$$Q(S, S^c) = \sum_{\substack{s\in S \\ w\in S^c}} \pi_s \mathbf{P}_{s,w} \geq \pi_u \mathbf{P}(u, v) \geq \pi_{\min}\alpha.$$

Since $1 - \beta < 1$, we finally obtain

$$\mathrm{Var}_{S_t}\left[\pi(S_{t+1})\right] \geq \beta \pi_{\min}^2 \alpha^2.$$

□

Finally, we derive an upper bound on the amount by which S_t can increase in one iteration.

Lemma 3.7.11. *For any time t and $S_0 = \{x\}$, $\pi(S_{t+1}) \leq \left(\frac{1-\beta}{\alpha} + 1\right) \frac{\pi_{\max}}{\pi_{\min}} \cdot \pi(S_t)$.*

Proof. Since

$$S_{t+1} = \left\{ y \in \Omega : \frac{\sum_{x \in S_t} \pi(x)\mathbf{P}(x,y)}{\pi(y)} \geq U \right\}.$$

If U decreases to 0, then every $y \in S_{t+1}$ is at least connected to an $x \in S_t$. In other words, $\mathbf{P}(x,y) > 0$ for $x \in S_t$ and $y \in S_{t+1}$. Hence $|S_{t+1}| \leq (\frac{1-\beta}{\alpha} + 1)|S_t|$.

We also know that

$$\pi(S_{t+1}) \leq |S_{t+1}| \cdot \pi_{\max} \leq \left(\frac{1-\beta}{\alpha} + 1 \right) \cdot |S_t| \cdot \pi_{\max} \leq \left(\frac{1-\beta}{\alpha} + 1 \right) \cdot \pi(S_t) \cdot \frac{\pi_{\max}}{\pi_{\min}}.$$

□

The proof of Theorem 3.5.3 follows then by combining Proposition 3.7.3, Lemma 3.7.4, Lemma 3.7.9, Lemma 3.7.10 and Lemma 3.7.11.

Proof of Theorem 3.5.3. With the help of the previous three lemmas, we can apply Proposition 3.7.3 with $M_0 = \pi(x)$, $\sigma \geq \beta^{1/2}\pi(x)\alpha$ and $D = \left(\frac{1-\beta}{\alpha}\right) \frac{\pi_{\max}}{\pi_{\min}}$ to obtain

$$\begin{aligned}
|\mathbf{P}^t(x,y) - \pi(y)| &\leq \frac{\pi(y)}{\pi(x)} \mathbb{P}_{\{x\}}[\tau > t] \\
&\leq \frac{\pi(y)}{\pi(x)} \frac{2\pi(x)}{\sigma} \sqrt{\frac{D}{t}} \\
&= \frac{2\pi(y)}{\beta^{1/2}\pi_{\min}\alpha} \sqrt{\frac{(\frac{1-\beta}{\alpha} + 1)\frac{\pi_{\max}}{\pi_{\min}}}{t}} \\
&\leq \frac{\pi_{\max}^{3/2}}{\pi_{\min}^{3/2}} \cdot \frac{2}{\beta^{1/2}\alpha} \sqrt{\frac{1-\beta+\alpha}{\alpha t}}.
\end{aligned}$$

□

3.7.2 Proof of Theorem 3.5.4

We show the upper bound for $|\mathbf{M}^t(u,v) - 1/n| = O(t^{-1/2})$. Here for random matching, we cannot use this bound because unlike balancing circuit this random matching model does not have a fixed transition matrix. Also it is not "irreducible".

The random matching model we analyse in this thesis is the model defined by Muthukrishnan and Ghosh. In the first stage, every node picks an incident edge independently with probability $O(1/d)$. In the second stage, we consider the matching formed by all edges that are not incident to any other edge chosen in the first stage.

3.7.2.1 Bounded degree case

We consider the easier case of a bounded-degree graph before moving on to the general case of a d-regular graph.

We will consider a random matching model with the property that each edge is included in any round with probability at least $p_{min} = \Omega(1/d)$, where d is the degree of the graph.

Lemma 3.7.12. *Let G be any connected, bounded-degree graph and $\mathbf{P}$ be the matrix representing a random matching. Let σ be any probability vector with $\|\sigma\|_2^2 = \beta$, then*

$$\mathbb{E}\left[\|\sigma\|_2^2 - \|\mathbf{P}\sigma\|_2^2\right] = \Omega(\beta^3).$$

We assume that $1/n \leq \beta \leq 1$ because β should not be too small otherwise it is almost balanced. (Also, if it is just $1/n$, then the lemma is not true since we cannot expect to have any improvement.)

Proof. First note that if $\sigma_{\max} \leq \beta/100$, then since $\|.\|_2^2 \leq \|.\|_1 \cdot \|.\|_\infty$ we obtain $\|\sigma\|_2^2 \leq \beta/100$, a contradiction. Hence we must have $\sigma_{\max} \geq \beta/100$. Consider now a shortest path from the vertex, say, i corresponding to $\sigma_{\max}$ to a vertex j with $\sigma_j \leq \beta/200$. Since σ is a probability vector, the length of the shortest path is at most $200/\beta$. Let us denote this shortest path by $(v_1 = i, v_2, \ldots, v_\ell = j)$, where $\ell \leq 200/\beta$.

Every edge, in particular any edge $\{v_k, v_k + 1\}$ with $1 \leq k < \ell$, is included in the random matching with probability $p_{\min} = \Omega(1/d) = \Omega(1)$, and if the edge is included no other edges incident to k or $k+1$ are included, which means that in this case the contribution to $\|\sigma\|_2^2 - \|\mathbf{P}\sigma\|_2^2$ will be exactly

$$\sigma_{v_k}^2 + \sigma_{v_{k+1}}^2 - 2 \cdot \left(\frac{\sigma_{v_k} + \sigma_{v_{k+1}}}{2}\right)^2 = \sigma_k^2/2 - \sigma_{v_k}\sigma_{v_{k+1}} + \sigma_{v_{k+1}}^2/2 = \frac{1}{2}(\sigma_{v_k} - \sigma_{v_{k+1}})^2.$$

By considering the (expected) contribution of all edges on the shortest path, we obtain

$$\begin{aligned}\mathbb{E}\left[\|\sigma\|_2^2 - \|\mathbf{P}\sigma\|_2^2\right] &\geq \sum_{k=1}^{\ell-1} \mathbb{P}\left[\{v_k, v_{k+1}\} \text{ is in the random matching}\right] \cdot \frac{1}{2}(\sigma_{v_k} - \sigma_{v_{k+1}})^2 \\ &\geq \frac{1}{2} p_{\min} \cdot \sum_{k=1}^{\ell-1} (\sigma_{v_k} - \sigma_{v_{k+1}})^2.\end{aligned}$$

Since $\ell \leq 200/\beta$, $\sigma_{v_1} \geq \beta/100$ and $\sigma_{v_\ell} \leq \beta/200$, it follows that

$$\begin{aligned}\mathbb{E}\left[\|\sigma\|_2^2 - \|\mathbf{P}\sigma\|_2^2\right] &\geq \frac{1}{2} p_{\min} \cdot 200/\beta \cdot \left(\frac{\sigma_{v_1} - \sigma_{v_\ell}}{200/\beta}\right)^2 \\ &\geq \Omega(1) \cdot \frac{(\beta/200)^2}{200/\beta} = \Omega(\beta^3),\end{aligned}$$

which completes the proof. □

3.7.2.2 Unbounded degree case

Let us now consider the case where G may be of unbounded degree. We first make the following simple observation:

Observation 3.7.13. *Let G be a d-regular graph. Then for any integer $x \geq 1$, the number of vertices reachable in x hops is at least $\Omega(d \cdot x)$.*

Lemma 3.7.14. *Let G be any connected graph and $\mathbf{P}$ be the matrix representing a random matching. Let σ be any probability vector with $\|\sigma\|_2^2 = \beta$, then*

$$\mathbb{E}\left[\|\sigma\|_2^2 - \|\mathbf{P}\sigma\|_2^2\right] = \Omega(\beta^3).$$

Proof. We proceed similarly as the proof before. However, using the lemma above we conclude that there must be a shortest path $(v_1 = i, v_2, \ldots, v_\ell = j)$ with $\sigma_i \geq \beta/100$ and $\sigma_j \leq \beta/200$, but here we have the stronger condition $\ell \leq 200/(\beta \cdot d)$. (This is because within ℓ hops, we reach $\Omega(d \cdot \ell)$ vertices and not all of these $\Omega(d \cdot \ell)$ vertices r can have $\sigma_r \geq \beta/200$, since σ is a probability vector.) Hence as before we conclude that

$$\mathbb{E}\left[\|\sigma\|_2^2 - \|\mathbf{P}\sigma\|_2^2\right] \geq \frac{1}{2} p_{\min} \cdot 200/(\beta d) \cdot \left(\frac{\sigma_{v_1} - \sigma_{v_\ell}}{200/(\beta d)}\right)^2 \geq \Omega(1/d) \cdot \frac{(\beta/200)^2}{200/(\beta d)} = \Omega(\beta^3).$$

□

Now we have proved the expected one-step decrease. We prove the original theorem by accumulating t one-step decreases.

Proof of Theorem 3.5.4. Now we bound the time used to reduce the norm from β to $\beta/2$ by bounding the expected drops across all rounds. Let $X_t = \|\sigma_t\|_2^2$, $t_{goal} := \{\min\ t : X_t \leq \beta/2\}$. We know that for any t, $\frac{1}{n} \leq \beta_t = \|\sigma_t\|_2^2 \leq 1$. Also we can prove that β_t is non-increasing and $\beta_t \geq 2^{-t}\beta$. We have

$$\mathbb{E}\left[X_{t+1} | \mathbb{1}_{\{X_t > \beta/2\}}\right] \leq X_t - X_t^3.$$

This one is a strong and general thing we want to prove, but it might be difficult. Hence it should suffice to have the following inequality:

$$\begin{aligned}
&\mathbb{E}\left[X_{t+1} | \mathbb{1}_{\{X_t > \beta/2\}}\right] \leq X_t - \beta^3/8, \\
&\mathbb{E}\left[X_{t+1} | \mathbb{1}_{\{t_{goal} \geq t+1\}}\right] \leq X_t - \beta^3/8, \\
&\mathbb{E}\left[X_{t+1}\mathbb{1}_{\{t_{goal} \geq t+1\}} | \mathbb{1}_{\{t_{goal} \geq t+1\}}\right] \leq (X_t - \beta^3/8)\mathbb{1}_{\{t_{goal} \geq t+1\}} \leq (X_t - \beta^3/8)\mathbb{1}_{\{t_{goal} \geq t\}}.
\end{aligned}$$

The last inequality holds because first if $t_{goal} < t$ or $t_{goal} \geq t+1$ then they are equal, second, if $t+1 > t_{goal}$ and $t = t_{goal}$ then it is also true. Take expectation on both sides,

we have

$$\begin{aligned}\mathbb{E}\left[X_{t+1}\mathbb{1}_{\{t_{goal}\geq t+1\}}\right] &\leq \mathbb{E}\left[X_t\mathbb{1}_{\{t_{goal}\geq t\}}\right] - \beta^3/8\cdot\mathbb{P}\left[t_{goal}\geq t\right]\\ &\leq \mathbb{E}\left[X_{t-1}\mathbb{1}_{\{t_{goal}\geq t-1\}}\right] - \beta^3/8\cdot\left(\mathbb{P}\left[t_{goal}\geq t-1\right]+\mathbb{P}\left[t_{goal}\geq t\right]\right)\\ &\leq ...\\ &\leq \mathbb{E}\left[X_0\mathbb{1}_{\{t_{goal}\geq 0\}}\right] - \beta^3/8\cdot\sum_{i=0}^{t}\mathbb{P}\left[t_{goal}\geq i\right].\end{aligned}$$

We take the limit on both sides.

$$\begin{aligned}\lim_{t\to\infty}\mathbb{E}\left[X_{t+1}\mathbb{1}_{\{t_{goal}\geq t+1\}}\right] &\leq \mathbb{E}\left[X_0\right] - \beta^3/8\cdot\sum_{i=0}^{\infty}\mathbb{P}\left[t_{goal}\geq i\right]\\ (\beta^3/8)\cdot\mathbb{E}\left[t_{goal}\right] &\leq \beta - \beta^3/8 - \lim_{t\to\infty}\mathbb{E}\left[X_{t+1}\mathbb{1}_{\{t_{goal}\geq t+1\}}\right]\\ &\leq \beta - \beta^3/8 - c\cdot\lim_{t\to\infty}\mathbb{P}\left[\mathbb{1}_{\{t_{goal}\geq t+1\}}\right].\end{aligned}$$

Now we show that $\lim_{t\to\infty}\mathbb{P}\left[\mathbb{1}_{\{t_{goal}\geq t+1\}}\right] = 0$, in other words, the process will eventually stop. To prove it can stop, we show that there exists small $\epsilon_1, \epsilon_2 > 0$,

$$\mathbb{P}\left[X_t - X_{t+1}\geq\epsilon_1\right]\geq\epsilon_2.$$

In random matching model, we know that whenever we match two nodes, our X_t would not increase. We define $Z = X_t - X_{t+1}$. Then we have $0\leq Z\leq 1$, $\mathbb{E}\left[Z\right] = \Omega(\beta^3/8)$, and $Z\geq 0$.

Theorem 3.7.15 (The reverse Markov inequality)**.** *Let X be a random variable that is never larger than B. Then, for all $a < B$,*

$$\mathbb{P}\left[X\leq a\right]\leq\frac{\mathbb{E}\left[B-X\right]}{B-a}.$$

We apply the above theorem on Z and we want to get $\mathbb{P}\left[X\leq a\right]\leq\frac{1}{\epsilon}$ for positive a,ϵ. Hence we choose $B = 1$ and $a = 1-\epsilon+\epsilon\cdot\mathbb{E}\left[Z\right]$. Then we let $\epsilon = 1 + c\cdot\mathbb{E}\left[Z\right]$ where c is a constant less than 1. Now we have

$$\mathbb{P}\left[Z\leq 1-(1+c\cdot\mathbb{E}\left[Z\right])+(1+c\cdot\mathbb{E}\left[Z\right])\cdot\mathbb{E}\left[Z\right]\right]\leq\frac{1}{1+c\cdot\mathbb{E}\left[Z\right]}.$$

By adjusting terms:

$$\mathbb{P}\left[Z\geq c\cdot\mathbb{E}\left[Z\right]\right]\geq\frac{1}{1+c\cdot\mathbb{E}\left[Z\right]}.$$

Hence now we have proved that there exist positive ϵ_1,ϵ_2 such that $\mathbb{P}\left[X_t - X_{t+1}\geq\epsilon_1\right]\geq\epsilon_2$. Therefore, the process will eventually stop and we get $\mathbb{E}\left[t_{goal}\right]\leq 8/\beta^2 - 1 = O(\beta^2)$.

We prove that based on the previous result. We can see that to reduce the $X_i = \frac{c}{\sqrt{t}}$

to its half. It only takes $\frac{t}{c^2} < t$ time. Hence, we can get a bound on the expected value of $X_t = \|\mathbf{M}^t(u,.)\|_2^2$, which is just $O(\frac{1}{\sqrt{t}})$. □

3.8 Proof of Theorem 3.5.6

We now prove the following discrepancy bound that depends on the $\lambda(\mathbf{M})$, as defined in Section 3.4.

Proof of Theorem 3.5.6. By [85, Lemma 2.4], for any pair of vertices $u, v \in V$,

$$\left|\mathbf{M}^t(u,v) - \frac{1}{n}\right| \leq \lambda(\mathbf{M})^{t/2}.$$

Hence by Lemma 3.7.1 $\|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2 = O(\lambda(\mathbf{M})^{t/4})$, and the bound on the discrepancy follows from Theorem 3.5.1 and the union bound over all vertices. □

3.9 Proof of Theorem 3.5.8

Proof of Theorem 3.5.8. Recall that when deriving general bounds, we had

$$\xi_t(u) = \sum_{w \in V} \xi_0(w)\mathbf{M}_t(w,u),$$

the load $\xi_t(u)$ is just a weighted sum of i.i.d. random variables and we obtain

$$\xi_t(u) - \xi_t(v) = \sum_{w \in V} \xi_0(w) \cdot \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right),$$

which is in fact still a weighted sum of n i.i.d. random variables. The expectation is

$$\begin{aligned}\mathbb{E}\left[\xi_t(u) - \xi_t(v)\right] &= \mathbb{E}\left[\sum_{w \in V} \xi_0(w) \cdot \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right)\right] \\ &= \sum_{w \in V} \left(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v)\right) \mathbb{E}\left[\xi_0(w)\right] = 0,\end{aligned}$$

where the last equality holds since $\mathbf{M}$ is doubly stochastic and the fact that $\xi_0(w)$ are identically distributed.

Now applying Hoeffding's inequality (Theorem 2.2.2), we obtain that

$$\mathbb{P}\left[|\xi_t(u) - \xi_t(v)| \geq \delta\right] \leq 2 \cdot \exp\left(\frac{-2\delta^2}{(\text{range}(\xi_0))^2 \cdot \|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2^2}\right).$$

This is what we may have when still applying Hoeffding. If we assume the initial

discrepancy is K, then we would get $\delta = K\sqrt{\log n} \cdot t^{-1/4}$. We should also use the following theorem, which is a classic tail bound.

Theorem 3.9.1 (Theorem 2.7, [63]). *Let the random variables $X_1, ..., X_n$ be independent, with $X_k - \mathbb{E}[X_k] \leq b$ for each k. Let $S_n = \sum X_k$, and let S_n have expected value μ and variance V (the sum of the variances of the X_k). Then for any $t \geq 0$,*

$$\begin{aligned}\mathbb{P}[S_n - \mu \geq t] &\leq e^{-(V/b^2)((1+\epsilon)\ln(1+\epsilon)-\epsilon)} \\ &\leq e^{-\frac{t^2}{2V(1+(bt/3V))}},\end{aligned}$$

where $\epsilon = bt/V$.

We attempt to bound the discrepancy between an arbitrary pair of nodes u, v. Here $X_k = \xi_k^{(0)} \cdot (\mathbf{M}^t(k,u) - \mathbf{M}^t(k,v))$, so $\mu = 0$, hence we have (replace t by δ):

$$\mathbb{P}[S_n \geq \delta] \leq e^{-\frac{\delta^2}{2V+(2b\delta/3)}}. \tag{3.6}$$

So here δ is a value greater than 0. $S_n = \sum(w)\xi_0(w)(\mathbf{M}^t(w,u) - \mathbf{M}^t(w,v))$, $V = var(S_n)$.

$$\mathrm{Var}[S_n] = \mathrm{Var}[\xi_0] \, \|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2^2.$$

Note that in our case, we have expectations for ξs but without variances. To deal with this infinite variance, we manually assume that after we've sampled loads for initial inputs the maximum input is K. The probability that this could happen is upper bounded by $K^{-\alpha}$. We can let $b = O(Kt^{-1/2} - \mathbb{E}[\xi(w)])$.

For a certain node w, let $c = \sum_{i=0}^{K} \mathbb{P}[\xi_0(w) = i]$.

$$\begin{aligned}\mathbb{P}[\xi_0(w) = i | \xi_0(w) \leq K] &= \frac{\mathbb{P}[\xi_0(w) = i]}{c}, \\ \mathbb{E}[\xi_0(w) | \xi_0(w) \leq K] &= \frac{1 \cdot \frac{1}{1^\alpha} + 2 \cdot \frac{1}{2^\alpha} + ... + K \cdot \frac{1}{K^\alpha}}{c} \\ &= \frac{\frac{1}{1^{\alpha-1}} + \frac{1}{2^{\alpha-1}} + ... + \frac{1}{K^{\alpha-1}}}{c}, \\ \mathrm{Var}[\xi_0(w) | \xi_0(w) \leq K] &= \mathbb{E}\left[\xi^2\right] - \mathbb{E}[\xi]^2 \\ &= \frac{\frac{1}{1^{\alpha-2}} + \frac{1}{2^{\alpha-2}} + ... + \frac{1}{K^{\alpha-2}}}{c} - \left(\frac{\frac{1}{1^{\alpha-1}} + \frac{1}{2^{\alpha-1}} + ... + \frac{1}{K^{\alpha-1}}}{c}\right)^2 \\ &\leq \frac{K^{(3-\alpha)}}{c} - \frac{K^{(4-2\alpha)}}{c^2}.\end{aligned}$$

Let $\mathcal{E}_K$ be the event that the maximum input load is K and disc_t is the discrepancy of the load vector at time t, then

$$\mathbb{P}[\mathrm{disc}_t \geq \delta \,|\, \mathcal{E}_K] = \frac{\mathbb{P}[\mathrm{disc}_t \geq \delta \cap \mathcal{E}_K]}{\mathbb{P}[\mathcal{E}_K]}. \tag{3.7}$$

In our case, we bound $\mathbb{P}[\text{disc} \geq \delta \,|\, \mathcal{E}_K]$ to be smaller than $1/n^2$. So

$$\text{Var}[\xi_0 | \forall w, \xi_0(w) \leq K] \leq K.$$

We substitute this in (3.6) above. Suppose we want the probability to be n^{-2}. We then know that

$$\frac{3\delta^2}{6K^{(3-\alpha)}\|.\|_2^2 + 2Kt^{-1/2} \cdot \delta} = 2\log(n). \tag{3.8}$$

Since $\delta > 1$, we know $6K^{(3-\alpha)}\|.\|_2^2 \leq 2Kt^{-1/2} \cdot \delta$. Hence, we can derive $\delta = O(\log(n)Kt^{-1/2})$. Formally, our conclusion would be: when knowing the initial discrepancy K, for an arbitrary pair u, v, we know their discrepancy satisfies:

$$\mathbb{P}[|\xi_t(u) - \xi_t(v)| \geq \delta \,|\, \mathcal{E}_K] \leq \frac{1}{n^2}$$

where $\delta = O(\log(n)Kt^{-1/2})$.

So, if we want to get a general bound on the discrepancy, without having this specific K, what we do is to find the 'expected discrepancy'. Let δ_K be the discrepancy when knowing a specific K. So for any pair u, v, their expected discrepancy in the end is

We try to find the expected maximum initial load. For every initial load x_i which is drawn from a power law distribution. We can prove that $\exists c_1, c_2 > 0, c_1 \cdot K^{1-\alpha} \leq \mathbb{P}[x_i \geq K] \leq c_2 \cdot K^{1-\alpha}$. By using this integral trick, we know that $\int_{K+1}^{\infty} x^{-\alpha} dx \leq \mathbb{P}[x_i \geq K] \leq \int_K^{\infty} x^{-\alpha} dx$. Hence we can actually be certain that $c_1 = \left(\frac{K}{K+1}\right)^{\alpha-1} \cdot \frac{1}{\alpha-1}$ and $c_2 = \frac{1}{\alpha-1}$ in the above formula.

$$\mathbb{P}\left[\max_u \xi_0(u) \leq K\right] = (\mathbb{P}[X \leq K])^n \leq (1 - c_1 \cdot K^{1-\alpha})^n \leq \exp(-c_1 \cdot n \cdot K^{1-\alpha}).$$

This probability is smaller than $\exp(-1/2) \approx 0.6$ if $K \leq n^{1/(\alpha-1)}$. If K becomes larger, this probability will be smaller. Therefore we can choose K to be a little larger, say, $K \geq c_3 \cdot n^{1/(\alpha-1)}$ where c_3 is a small but sufficiently large constant such that the probability above is less than $1/2$. Hence with this K, we have

$$\mathbb{P}[\text{disc}_0 \geq K] \geq 1 - \exp(-c_1 \cdot n \cdot K^{1-\alpha}) \geq 1/2.$$

Therefore we have got a lower bound $\mathbb{E}[\text{disc}_0^{\max}] = \Omega(n^{1/(\alpha-1)})$. Now we want to upper bound the discrepancy.

$$\mathbb{P}\left[\max_u \xi_0(u) \leq K\right] \geq (1 - c_2 \cdot K^{1-\alpha})^n \geq 1 - c_2 \cdot n \cdot K^{1-\alpha} \geq 10^{-c_2 \cdot n \cdot K^{1-\alpha}}.$$

The last inequality holds since the second last formula is linear and the last one is expo-

nentially decreasing. More generally, we have for any integer $\tau \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}\left[\tau \cdot n^{1/(\alpha-1)} \leq \mathrm{disc}_0^{\max} \leq 2 \cdot \tau \cdot n^{1/(\alpha-1)}\right] &\leq \mathbb{P}\left[\mathrm{disc}_0^{\max} \geq \tau \cdot n^{1/(\alpha-1)}\right] \\ &\leq 1 - 10^{-c_2 \cdot \tau^{1-\alpha}} \\ &= 1 - \exp(-c_2 \ln(10) \cdot \tau^{1-\alpha}) \\ &\approx 1 - (1 - c_2 \ln(10) \cdot \tau^{1-\alpha}) \\ &\approx c_2 \ln(10) \cdot \tau^{1-\alpha},\end{aligned}$$

where we have used the fact that $\exp(-x) \approx 1 - x$ for any x close to 0. Hence the upper bound for the expected value of $\max_u \xi(u)$ by using buckets.

$$\begin{aligned}\mathbb{E}\left[\max_u \xi_0(u)\right] &= \sum_{K=1}^{\infty} \mathbb{P}\left[\max_u \xi_0(u) \geq K\right] \\ &\leq \mathbb{P}\left[1 \leq \max_u \xi_0(u) \leq n^{1/(\alpha-1)}\right] \cdot n^{1/(\alpha-1)} \\ &\quad + \sum_{\delta=0}^{\infty} \mathbb{P}\left[2^{\delta} \cdot n^{1/(\alpha-1)} \leq \max_u \xi_0(u) \leq 2^{\delta+1} \cdot n^{1/(\alpha-1)}\right] \cdot 2^{\delta+1} \cdot n^{1/(\alpha-1)} \\ &\leq n^{1/(\alpha-1)} + 2c_2 \ln(10) n^{1/(\alpha-1)} \cdot \sum_{\delta=0}^{\infty} 2^{\delta(2-\alpha)} \\ &= O(n^{1/(\alpha-1)}).\end{aligned}$$

Till now we have proved that $\mathbb{E}\left[\max_u \xi_0(u)\right] = \Theta(n^{1/(\alpha-1)})$. By applying our Theorem (3.6), we can say that the expected discrepancy is $O(\log(n) n^{1/(\alpha-1)} t^{-1/2})$.

□

3.10 Applications to different graph topologies

3.10.1 Cycles

Recall that for the cycle, $V = \{0, \ldots, n-1\}$ is the set of vertices, and the distance between two vertices is $\mathrm{dist}(x, y) = \min\{y - x, x + n - y\}$ for any pair of vertices $x < y$.

The upper bound on the discrepancy follows directly from Corollary 3.5.5, and it only remains to prove the lower bound. To this end, we will apply the lower bound in Theorem 3.5.1 and need to derive a lower bound on $\|\mathbf{M}^t(., u) - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2$. Intuitively, if we had a simple random walk, we could immediately infer that this quantity is $\Omega(1/\sqrt{t})$. Since after t steps, the random walk is with probability $\approx 1/\sqrt{t}$ at any vertex with distance at most $O(\sqrt{t})$. To prove that this also holds for the load balancing process, we first derive a concentration inequality that upper bounds the probability for the random walk to reach a distant state:

Lemma 3.10.1. *Consider the standard balancing circuit model on the cycle with round matrix* $\mathbf{M}$*. Then for any* $u \in V$ *and* $\delta \in (0, n/2-1)$*, we have*

$$\sum_{v \in V : \mathrm{dist}(u,v) \geq \delta} \mathbf{M}^t(u,v) \leq 2 \cdot \exp\left(-\frac{(\delta-2)^2}{8t}\right).$$

Proof. The proof of the lemma above makes uses of the following variant of Azuma's concentration inequality for martingales, which can be for instance found in McDiarmid's survey on concentration inequalities.

Lemma 3.10.2 ([63, Theorem 3.13 & Inequality 41])**.** *Let* $Z_1, Z_2, \ldots, Z_n$ *be a martingale difference sequence with* $a_k \leq Z_k \leq b_k$ *for each* k*, for suitable constants* a_k, b_k*. Then for any* $\delta \geq 0$*,*

$$\mathbb{P}\left[\max_{1 \leq j \leq n} \left|\sum_{i=1}^{j} Z_i\right| > \delta\right] \leq 2 \cdot \exp\left(-\frac{2\delta^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right) \tag{3.9}$$

Note that the balancing circuit on the cycle corresponds to the following random walk $(X_1, X_2, \ldots, X_t)$ on the vertex set $V = \{-n/2+1, \ldots, 0, \ldots, n/2-1\}$, where for any time-step $t \in \mathbb{N}$, X_t denotes the position of the random walk after step t. First, we consider the transition for any odd s: If X_s is odd, then with probability $1/2$, $X_{s+1} = X_s + 1$ and otherwise $X_{s+1} = X_s$. If X_s is even, then with probability $1/2$, $X_{s+1} = X_s - 1$ and otherwise $X_{s+1} = X_s$ (additions and subtractions are under the implicit assumptions that $-n/2+1 \equiv n/2-1$ and $n/2 \equiv -n/2+1$). The case for even s is analogous.

We will couple the random walk $(X_t)_{t \geq 0}$ with another random walk $(Y_t)_{t \geq 0}$ on the integers $\mathbb{N}$, where again Y_t denotes the position of the walk after step t. The transition probabilities are exactly the same as for the walk $(X_t)_{t \geq 0}$, the only difference is that we don't use the equivalences $-n/2+1 \equiv n/2-1$ and $n/2 \equiv -n/2+1$. It is clear that we can couple the transitions of the two walks so that they evolve identically as long as the walks do not reach any of the two boundary points $-n/2+1$ or $n/2-1$.

Let us first analyze $\mathbb{E}[Y_t]$ for an odd time step. As described above, the distribution of $Y_t - Y_{t-1}$ depends on whether Y_{t-1} is even or not. However, notice regardless of where the random walk is at step $t-2$, the random walk will be at an odd or even vertex at step $t-1$ with probability $1/2$ each. Hence for any starting position y,

$$\begin{aligned}
&\mathbb{E}[Y_t - Y_{t-1} \mid Y_0 = y] \\
&= \mathbb{P}[Y_{t-1} \text{ even}] \cdot \left(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0\right) + \mathbb{P}[Y_{t-1} \text{ odd}] \cdot \left(\frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot 0\right) = 0,
\end{aligned}$$

and further,

$$|Y_1 - Y_0| \leq 1.$$

Combining the last two inequalities shows that for any start vertex y,

$$|\mathbb{E}[Y_t \mid Y_0 = y] - y| \leq 1.$$

With the same arguments as before we conclude that for any fixed start vertex $Y_0 = y_0$,

$$\max_{a,b \in V} |\mathbb{E}[Y_t - Y_{t-1} \mid Y_1 = a] - \mathbb{E}[Y_t - Y_{t-1} \mid Y_1 = b]| \leq 2,$$

because the expected differences of $Y_t - Y_{t-1}$ are all zero whenever $t \geq 3$.

Let us now consider the martingale $W_i = \mathbb{E}[Y_t \mid Y_0, Y_1, \ldots, Y_i]$, and let $Z_i := W_i - W_{i-1}$ be the corresponding martingale difference sequence. As shown before, $|W_i - W_{i-1}| \leq 2$. Hence by Lemma 3.10.2,

$$\mathbb{P}\left[\max_{1 \leq j \leq t} \left|\sum_{i=1}^{j} Z_i\right| \geq \delta\right] \leq 2 \cdot \exp\left(-\frac{\delta^2}{8 \cdot t}\right).$$

If for every $1 \leq j \leq t$, $\sum_{i=1}^{j} W_i < \delta$ holds, then this implies both random walks $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ behave identically since none of them ever reaches any of the two boundary points $-n/2 + 1$ or $n/2 - 1$. In particular we conclude that for the original walk $(X_t)_{t \geq 0}$,

$$\begin{aligned}
\sum_{v \in V : \mathrm{dist}(u,v) \geq \delta} \mathbf{M}^t(u, v) &= \mathbb{P}\left[\left|\sum_{i=1}^{t} X_i\right| \geq \delta\right] \\
&\leq \mathbb{P}\left[\max_{1 \leq j \leq t} \left|\sum_{i=1}^{j} X_t\right| \geq \delta\right] \\
&= \mathbb{P}\left[\max_{1 \leq j \leq t} \left|\sum_{i=1}^{j} Y_t\right| \geq \delta\right] \\
&\leq \mathbb{P}\left[\max_{1 \leq j \leq t} \left|\sum_{i=1}^{j} Z_t\right| \geq \delta - 2\right] \leq 2 \cdot \exp\left(-\frac{(\delta - 2)^2}{8 \cdot t}\right),
\end{aligned}$$

where the second-to-last inequality is due to the fact that $\mathbb{E}\left[\left|\sum_{i=1}^{j} Y_t\right|\right] \leq 2$. □

With the help Lemma 3.10.1, we can indeed verify our intuition:

Lemma 3.10.3. *Consider the standard balancing circuit model on the cycle with round matrix* $\mathbf{M}$. *Then for any vertex* $u \in V$, $\|\mathbf{M}^t(., u) - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2 = \Omega(1/\sqrt{t})$.

Proof. Define $S_\delta := \{w \in V : \mathrm{dist}(w, u) \leq \delta\}$, so that $|S_\delta| = 2\delta$. With $\delta = 20\sqrt{t}$ and $t \geq 10$,

$$\sum_{w \in S_\delta} \mathbf{M}^t(w, u) = 1 - \sum_{w \notin S_\delta} \mathbf{M}^t(w, u) \geq 1 - 2 \cdot \exp\left(-\frac{(\delta - 2)^2}{8t}\right) \geq \frac{1}{2}.$$

By Cauchy-Schwarz inequality,

$$\|\mathbf{M}^t(w,\cdot)\|_2^2 \geq \sum_{w\in S_\delta} \left(\mathbf{M}^t(w,u)\right)^2 \geq \frac{1}{2\delta}\left(\sum_{w\in S_\delta} \mathbf{M}^t(w,u)\right)^2 \geq \frac{1}{2\delta}\left(\frac{1}{2}\right)^2 = \Omega(t^{-1/2}).$$

□

Lemma 3.10.3 also proves that the factor $\sqrt{1/t}$ in the upper bound in Theorem 3.5.3 is the best possible. The lower bound on the discrepancy now follows by combining Lemma 3.10.3 with Theorem 3.5.1 and Lemma 3.7.1 stating that for any vertex $u \in V$, there exists another vertex $v \in V$ such that $\|\mathbf{M}^t(.,u) - \mathbf{M}^t(.,v)\|_2^2 \geq \|\mathbf{M}^t(.,u) - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2 = \Omega(1/\sqrt{t})$.

3.10.2 Tori

In this section we consider r-dimensional tori, where $r \geq 1$ is any constant. For the upper bound, note that the computation of $\mathbf{M}^t_{.,.}$ can be decomposed to independent computations in the r dimensions, and each dimension has the same distribution as the cycle on $n^{1/r}$ vertices. Specifically, if we denote by $\widetilde{\mathbf{M}}$ the round matrix of the standard balancing circuit scheme on the cycle with $n^{1/r}$ vertices and $\mathbf{M}$ is the round matrix of the r-dimensional torus with n vertices, then for any pair of vertices $x = (x_1,\ldots,x_r), v = (y_1,\ldots,y_r)$ on the torus we have $\mathbf{M}^t(x,y) = \prod_{i=1}^r \widetilde{\mathbf{M}}^t(x_i,y_i)$. From Theorem 3.5.3, $|\widetilde{\mathbf{M}}^t(x_i,y_i) - \frac{1}{n^{1/r}}| = O(t^{-1/2})$, and therefore, since r is a constant,

$$\mathbf{M}^t(x,y) \leq \prod_{i=1}^r \left(\frac{1}{n^{1/r}} + O(t^{-1/2})\right) = O(t^{-r/2} + n^{-1}),$$

and thus $\left\|\mathbf{M}(x,y)^t - \frac{\mathbf{1}}{\mathbf{n}}\right\|_2^2 = O(t^{-r/2})$ for any pair of vertices x, y. Hence by Lemma 3.7.1, $\left\|\mathbf{M}^t(.,u) - \mathbf{M}^t_{.,v}\right\|_2^2 = O(t^{-r/2})$. Plugging this bound into Theorem 3.5.1 yields that the load difference between any pair of the nodes u and v at round t is at most $O(t^{-r/4}\cdot\sigma\cdot\log^{3/2} n + \sqrt{\log n})$ with probability at least $1 - 4n^{-1}$. The bound on the discrepancy now simply follows by the union bound.

We now turn to the lower bound on the discrepancy. With the same derivation as in Lemma 3.10.3 we obtain the following result:

Lemma 3.10.4. *Consider the standard balancing circuit model on the r-dimensional torus with round matrix $\mathbf{M}$. Then for any vertex $u \in V$, $\|\mathbf{M}^t(.,u) - \frac{\mathbf{1}}{\mathbf{n}}\|_2^2 = \Omega(t^{-r/2})$.*

The lower bound on the torus follows by combining Lemma 3.10.4 and Theorem 3.5.1.

3.10.3 Expanders

The upper bound $O(\lambda(\mathbf{M})^{t/4} \cdot \sigma \cdot (\log n)^{3/2} + \sqrt{\log n})$ for expanders follows immediately from Theorem 3.5.6. For the lower bound, since the round matrix consists of d matchings, it is easy to verify that whenever $\mathbf{M}^t(u,v) > 0$, we have $\mathbf{M}^t(u,v) \geq 2^{-d \cdot t}$. Consequently, for any vertex $u \in V$, $\left\|\mathbf{M}^t(.,u) - \frac{\mathbf{1}}{\mathbf{n}}\right\|_2^2 = \Omega(2^{-d \cdot t})$. Plugging this into Theorem 3.5.1 yields a lower bound on the discrepancy which is $\Omega(2^{-d \cdot t/2} \cdot \sigma / \sqrt{\log \sigma})$.

3.10.4 Hypercubes

For the hypercube, there is a worst-case bound of $\log_2 \log_2 n + O(1)$ [62, Theorem 5.1 & 5.3] for any input after $\log_2 n$ iterations of the dimension-exchange, i.e., after one execution of the round matrix. Hence, we will only analyze the discrepancy after s matchings, where $1 \leq s < \log_2 n$. By applying the same analysis as in Theorem 3.5.6, but now with $|\prod_{s=1}^{t} \mathbf{M}^{(s)}(u,v) - \frac{1}{n}| \leq 2^{-t}$, we obtain that the discrepancy is $O(2^{-t/2} \cdot \sigma \cdot (\log n)^{3/2} + \sqrt{\log n})$. Applying Theorem 3.5.1, we obtain the lower bound $\Omega(2^{-t/2} \cdot \sigma / \sqrt{\log \sigma})$.

3.11 Discussion

In this section, we discuss where our result fits in the literature. To see this, we need to compare the worst-case (classical settings) and the average-case (our setting). Our main result, as shown above, considers an average-case scenario where the load inputs are drawn from a fixed probability distribution, complementing previous worst-case analyses. To be more specific, for different topologies of networks such as cycles, tori, hypercubes and expanders, we obtain almost matching upper and lower bounds on the discrepancy, the difference between the maximum and the minimum load. Our bounds hold for a variety of probability distributions including uniform and binomial distributions but also distributions with unbounded range such as geometric, Poisson and power law distributions. For graphs with slow convergence like cycles and tori, our results demonstrate a substantial difference between the convergence in the worst-case and average-case.

An important ingredient in our analysis is a new upper bound on the t-step transition probability of a general Markov chain, which is derived by invoking the evolving set process.

3.11.1 Worst-case vs. Average-case

We will now compare our average-case to a worst-case scenario on cycles, 2D-tori and hypercubes. For the sake of concreteness, we always assume that the input is drawn from the uniform distribution $\mathsf{Uni}[0, 2K]$, where K will be specified later. Note that the total number of tokens is $\approx n \cdot K$, and the initial discrepancy will be $\Theta(K)$. Our choice for

the worst-case load vector will have the same number of tokens and initial discrepancy, however, the exact definition of the vector as well as the choice of the parameter K will depend on the underlying topology.

Cycles. As one representative of a worst-case setting, fix an arbitrary node $u \in V$ and let all nodes with distance at most $n/4$ initially have a load of $2K$ while all other nodes have load 0. This gives rise to a load vector with $n \cdot K$ tokens and initial discrepancy $2K$.

2D-Tori. Again, we fix an arbitrary node $u \in V$ and assign a load of $2K$ to the $n/2$-nearest neighbours of u and load 0 to the other nodes. Again, this defines a load vector with $n \cdot K$ tokens and initial discrepancy $2K$.

The next result provides a lower bound on the discrepancy for cycles and 2D-tori in the aforementioned worst-case setting. It essentially shows that for worst-case inputs, $\Omega(n^2)$ rounds and $\Omega(n)$ rounds are necessary for the cycle, 2D-tori, respectively, in order to reduce the discrepancy by more than a constant factor. This stands in sharp contrast to Corollary 3.5.5, proving a decay of the discrepancy by $\approx t^{-1/4}$, starting from the first round.

Proposition 3.11.1. *For the aforementioned worst-case setting on the cycle, it holds for any round $t > 0$ that* $\operatorname{disc}(x_t) \geq \frac{1}{8} \cdot K \cdot \left(1 - \exp\left(-\frac{n^2}{2048t}\right)\right) - \sqrt{64 \log n}$, *with probability at least $1 - n^{-1}$. Further, for 2D-tori, it holds for any round $t > 0$ that* $\operatorname{disc}(x_t) \geq \frac{1}{8} \cdot K \cdot \left(1 - \exp\left(-\frac{n}{2048t}\right)\right) - \sqrt{64 \log n}$, *with probability at least $1 - n^{-1}$.*

Proof. We first consider the case of a cycle. Let S_1 be the subset of nodes that have a non-zero initial load; so $|S_1| = n/2$. Clearly, there is a subset of nodes $S_2 \subseteq V$ with $|S_2| = n/8$ so that for each node $u \in S_2$, only nodes v with $\operatorname{dist}(u, v) \geq n/16$ can have $x_v^{(0)} > 0$.

We will now derive a lower bound on the discrepancy in this worst-case setting by upper bounding the load of vertices in the subset S_2. To lower bound the discrepancy at round t, recall that by Lemma 3.10.1 we have that

$$\sum_{v \in V : \operatorname{dist}(u,v) \geq \delta} \mathbf{M}^t(u, v) \leq 2 \cdot \exp\left(-\frac{(\delta - 2)^2}{8t}\right).$$

Let us now choose $\delta = n/16$, and we thus conclude that

$$\sum_{u \in S_1} \sum_{v \in S_2} \mathbf{M}^t(u, v) \leq \sum_{u \in S_1} \sum_{v \in V : \operatorname{dist}(u,v) \geq \delta} \mathbf{M}^t(u, v) \leq 2 \cdot |S_1| \cdot \exp\left(-\frac{n^2}{2048t}\right).$$

This implies for the total load of vertices in S_2 at time t:

$$\sum_{v \in S_2} \xi_t(v) = \sum_{v \in S_2} \sum_{u \in S_1} \xi_0(u) \cdot \mathbf{M}^t(u, v)$$

$$= 2K \sum_{u \in S_1} \sum_{v \in S_2} \mathbf{M}^t(u,v)$$
$$\leq K \cdot n \cdot \exp\left(-\frac{n^2}{2048t}\right),$$

where K is the average load. Recalling that $|S_2| = n/8$, by the pigeonhole principle there exists a node $v \in S_2$ such that

$$\xi_t(v) \geq \frac{1}{|S_2|} \cdot K \cdot n \cdot \exp\left(-\frac{n^2}{2048t}\right) \geq \frac{1}{8} \cdot K \cdot \exp\left(-\frac{n^2}{2048t}\right).$$

This immediately implies the following lower bound on the discrepancy:

$$\mathrm{disc}(\xi_t) \geq \bar{\xi} - \xi_t(v) \geq \bar{\xi} \cdot \left(1 - \exp\left(-\frac{n^2}{2048t}\right)\right),$$

where $\bar{\xi} = K$ is the average load. The corresponding lower bound on $\mathrm{disc}(x_t)$ follows by Theorem 3.6.1 and the union bound.

The proof for the 2-dimensional torus is almost identical. Again, let S_1 be the set of nodes that have a non-zero load. Clearly, there is a subset $S_2 \subseteq V$ with $|S_2| = n/8$ so that for each node $u \in S_2$, only nodes v with $\mathrm{dist}(u,v) \geq \sqrt{n}/16$ can have $x_v^{(0)} > 0$.

Let us now view $\mathbf{M}$ as the transition matrix of a Markov chain. Then $\mathbf{M}^t$ is obtained by running two independent Markov chains (one for each dimension), where each of the two Markov chains corresponds to the round matrix of the cycle. We can still apply Lemma 3.10.1 as before, even though here the size of each cycle is $\sqrt{n}$, to obtain that

$$\sum_{v \in V : \mathrm{dist}(u,v) \geq \delta} \mathbf{M}^t(u,v) \leq 2 \cdot \exp\left(-\frac{(\delta-2)^2}{8t}\right).$$

Here we choose $\delta = \sqrt{n}/16$, and the remaining part of the proof is exactly the same as before. □

Hypercube. Regarding the hypercube, we will consider only $\log_2 n$ rounds, since the discrepancy is $\log\log_2 n + O(1)$ after $\log_2 n$ rounds and $O(1)$ after $2\log_2 n$ rounds [62] in the worst case. A natural corresponding worst-case distribution is to have load $2K$ on all nodes whose $\log_2 n$-th bit is equal to 1 and load 0 otherwise. This way, the discrepancy is only reduced in the final round $\log_2 n$. However in the average case, the discrepancy reduces in a more smooth way, i.e., it keeps improving during the whole process.

3.12 Conclusion

Now we conclude our results and emphasise their importance. The main contribution of our load balancing study focuses on the analysis of the average-case load balancing. To

be more specific, this setting assumes the initial loads are sampled from a probability distribution instead of being chosen by an adversary as the worst-case analyses in the literature. On our way to the analysis, we develope a general Markov chain related theorem (Theorem 3.5.3), which is potentially useful elsewhere.

In particular, our main analysis consists of almost matching upper and lower bounds on the discrepancy, which are close to the classical results. In addition, we also establish a connection between the time used and the discrepancy, which is not often seen in previous works. We also compare our bounds and previous results on different graph topologies to see how our work fits in the field. In the next section, we will present our experimental results to simulate our regimes and validate our theorems.

3.13 Experiments

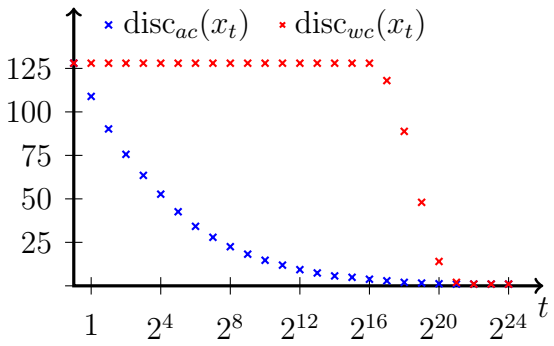
Lastly, we demonstrate our experimental results to simulate the process of our regimes. For each of the three graphs cycles, 2D-tori and hypercube, we consider two comparative experiments with an average-case load vector and a worst-case initial load vector each. The plots and tables on the next two pages display the results, where for each case we took the average discrepancy over **10 independent runs**.

The first experiment considers a "lightly loaded case", where the theoretical results suggest that a small (i.e., constant or logarithmic) discrepancy is reached well before the expected "worst-case load balancing times", which are $\approx n^2$ for cycles and $\approx n$ for 2D-tori. The second experiments consider a "heavily loaded case", where the theoretical results suggest that a small discrepancy is not reached faster than in the worst-case.

Specifically, for cycles and 2D-tori, we choose for the lightly loaded case $K = \sqrt{n}$ and for the heavily loaded case $K = n^2$. The experiments confirm the theoretical results in the sense that for both choices of K, we have a much quicker convergence of the discrepancy than in the corresponding worst cases. However, the experiments also demonstrate that only in the lightly loaded case we reach a small discrepancy quickly, whereas in the heavily loaded case there is no big difference between worst-case and average-case if it comes to the time to reach a small discrepancy.

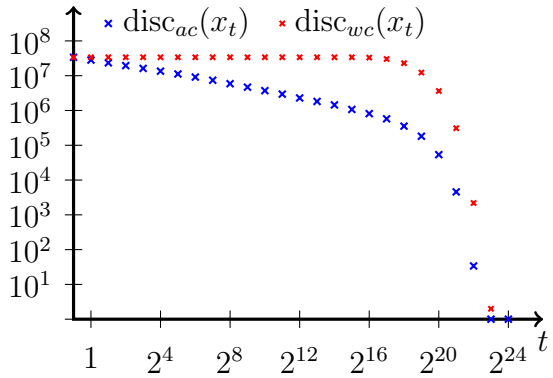
On the hypercube, since we are interested in the case where $1 \leq t \leq \log_2 n$, our bounds on the discrepancy indicates that we should choose K smaller than in the case of cycles and 2D-tori. That is why we choose $K = n^{1/4}$ in the lightly loaded case and $K = n$ in the heavily loaded case (As a side remark, we note that due to the symmetry of the hypercube, any initial load vector sampled from $\mathsf{Uni}[0, \beta \cdot (n-1)]$ is equivalent to an initial load vector sampled from $\mathsf{Uni}[0, n-1]$.) With these adjustments of K in both cases, the experimental results of the hypercube are inline with the ones for the cycle and 2D-tori.

The details of the experiments containing plots and tables with the sampled discrepancies can be found on the following pages.

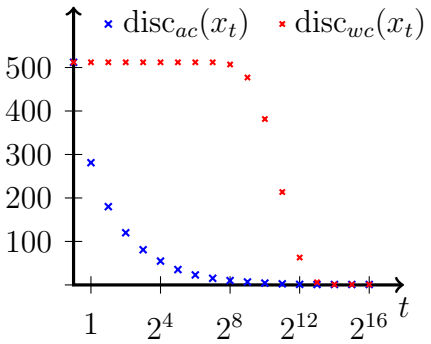


t	disc$_{ac}$	disc$_{wc}$
0	128.0	128.0
2^0	108.9	128.0
2^2	75.6	128.0
2^4	52.7	128.0
2^6	34.2	128.0
2^8	22.5	128.0
2^{10}	14.7	128.0
2^{12}	9.3	128.0
2^{14}	5.7	128.0
2^{16}	3.8	128.0
2^{18}	2.0	88.0
2^{20}	1.2	14.0
2^{22}	1.0	1.0
2^{24}	1.0	1.0

Experimental Results: Experiments (i) on the cycle with $n = 2^{12}$ and initial discrepancy $2^7 = 128$. Each data point is the average result of 10 independent runs.

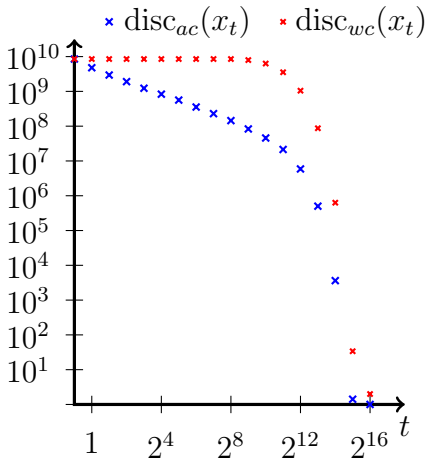


t	disc$_{ac}$	disc$_{wc}$
0	3.35×10^7	3.35×10^7
2^0	2.83×10^7	3.35×10^7
2^2	1.96×10^7	3.35×10^7
2^4	1.34×10^7	3.35×10^7
2^6	9.17×10^6	3.35×10^7
2^8	5.94×10^6	3.35×10^7
2^{10}	3.72×10^6	3.35×10^7
2^{12}	2.30×10^6	3.35×10^7
2^{14}	1.43×10^6	3.35×10^7
2^{16}	8.19×10^5	3.32×10^7
2^{18}	3.58×10^5	2.30×10^7
2^{20}	5.34×10^4	3.62×10^6
2^{22}	3.35×10^1	2.21×10^3
2^{24}	1.0	1.0

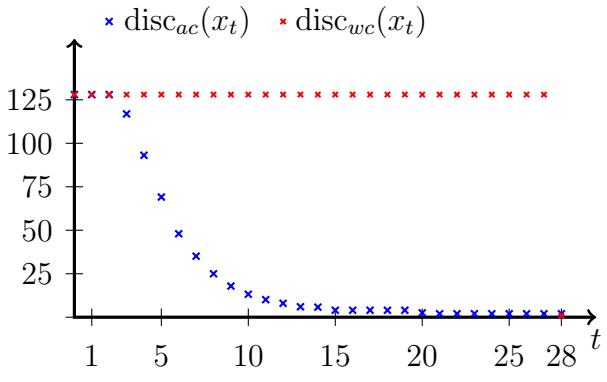


t	disc$_{ac}$	disc$_{wc}$
0	512.0	512.0
2^0	281.1	512.0
2^2	120.0	512.0
2^4	54.8	512.0
2^6	23.0	512.0
2^8	10.0	507.0
2^{10}	4.2	381.5
2^{12}	1.3	63.0
2^{14}	1.0	1.0
2^{16}	1.0	1.0

Experimental Results: Experiments (ii) on the cycle with $n = 2^{12}$ and initial discrepancy $2^{25} = 33,554,432$, and (iii) on the 2D-torus with $n = 2^{16}$ and initial discrepancy of 2^9. For the heavily loaded case, we used logarithmic scaling on the y-axis to highlight the behaviour when t is close to the worst-case load balancing time. Each data point is the average result of 10 independent runs.

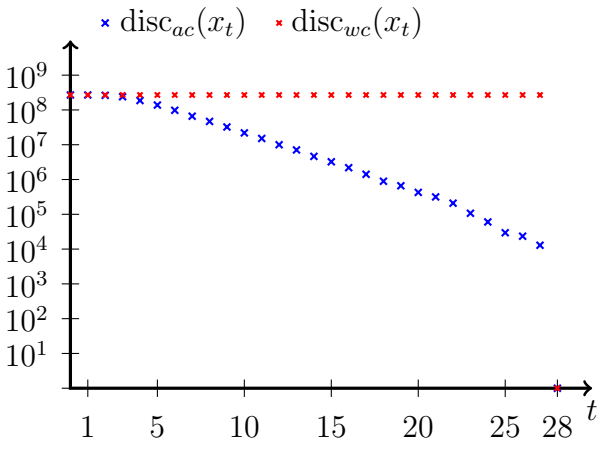


t	disc_{ac}	disc_{wc}
0	8.59×10^9	8.59×10^9
2^0	4.83×10^9	8.59×10^9
2^2	1.89×10^9	8.59×10^9
2^4	8.28×10^8	8.59×10^9
2^6	3.58×10^8	8.59×10^9
2^8	1.44×10^8	8.50×10^9
2^{10}	4.52×10^7	6.38×10^9
2^{12}	5.91×10^6	1.03×10^9
2^{14}	3.59×10^3	6.33×10^5
2^{16}	1.0	2.0



t	disc_{ac}	disc_{wc}
0	128.0	128.0
1	128.0	128.0
2	128.0	128.0
3	116.9	128.0
4	93.1	128.0
5	69.1	128.0
10	13.2	128.0
15	5.8	128.0
20	2.5	128.0
25	2.0	128.0
27	2.0	128.0
28	2.0	2.0

Experimental Results (cntd.): Experiments (iv) on the 2D-torus with $n = 2^{16}$ and initial discrepancy $2^{33} = 8,589,934,592$, (v) on the hypercube with $n = 2^{28}$ and initial discrepancy 256. For the heavily loaded cases, we used logarithmic scaling ($\log_2$) on the y-axis to highlight the behaviour when t is close to the worst-case load balancing time. Each data point is the average result of 10 independent runs.



t	disc_{ac}	disc_{wc}
0	2.68×10^8	2.68×10^8
1	2.68×10^8	2.68×10^8
2	2.65×10^8	2.68×10^8
3	2.41×10^8	2.68×10^8
4	1.88×10^8	2.68×10^8
5	1.37×10^8	2.68×10^8
10	2.21×10^7	2.68×10^8
15	3.26×10^6	2.68×10^8
20	4.23×10^5	2.68×10^8
25	2.98×10^4	2.68×10^8
27	1.29×10^4	2.68×10^8
28	3.0	3.0

Experimental Results (cntd.): Experiments (iv) on the 2D-torus with $n = 2^{16}$ and initial discrepancy $2^{33} = 8,589,934,592$, (v) on the hypercube with $n = 2^{28}$ and initial discrepancy 256, and (vi) on the hypercube with $n = 2^{28}$ and initial discrepancy of $2^{28} = 268,435,456$. For the heavily loaded cases, we used logarithmic scaling on the y-axis to highlight the behaviour when t is close to the worst-case load balancing time. Each data point is the average result of 10 independent runs.

Chapter 4

Random Walks on Dynamic Graphs

4.1 Introduction of random walks

4.1.1 Classical random walks

A random walk on a network is: given an undirected graph $G = (V, E)$, the walk starts at a fixed vertex. Then at each step, the random walk moves to a randomly chosen neighbour [1]. Due to their simplicity and locality, random walks are a very useful algorithmic primitive, especially in the design of distributed algorithms [84, 52, 29, 24, 82, 38, 30, 31, 53, 21]. In contrast to topology-driven algorithms, algorithms based on random walks benefit from a strong robustness against structural changes in the network.

One key property of random walks is that, under mild assumptions on the underlying network, they converge to a stationary distribution – an equilibrium state in which every vertex is visited proportionally often to its degree. The time for this convergence to happen is called *mixing time*, and understanding this time is crucial for many sampling or exploration related tasks. In particular, whenever a graph has a small *mixing time*, also its *cover time* (the expected time to visit all vertices of the graph) is small as well.

4.1.2 Motivation

While most of the classical research has focused on static graphs, many networks today are subject to dramatic changes over time. Hence understanding the theoretical power and limitations of *dynamic* graphs has been identified as one of the key challenges in computer science [64]. Several recent works have indeed considered this problem and studied the behavior of random walks [10, 11, 83, 36, 56, 76, 78] or similar processes [15, 25, 28, 44, 55] on such dynamic graphs, and their applications to distributed computing [10, 83, 55]. Moreover, rather than a property of the underlying network itself, dynamic graphs may

[1] In case of a *lazy* random walk, the walk would remain at the current location with probability 1/2, and otherwise move to a neighbour chosen uniformly at random.

naturally arise in distributed algorithms when communication is performed on a changing, possibly disconnected, subgraph like a spanning-tree or a matching (see, e.g., [22]).

4.1.3 Our model

In this chapter, we study the popular *evolving graph model*. That is, we consider sequences of graphs $G_1, G_2, \ldots$ over the same set of vertices but with a varying set of edges. This model has been studied in, for example, [11, 86, 56]. Both [11] and later [86] proved a collection of positive and negative results about the mixing time (and related parameters), and they assume a worst-case scenario where the changes to the graph are dictated by an oblivious, non-adaptive adversary. For example, [11] proved the following remarkable dichotomy. First, even if all graphs $G_1, G_2, \ldots$ are connected, small (but adversarial) changes to the stationary distribution can cause exponential mixing (and hitting) times. Secondly, if the sequence of connected graphs share the same stationary distribution i.e., the degrees (or relative degrees) of vertices are time-invariant, then mixing and hitting times are polynomial. This assumption about a time-invariant stationary distribution is crucial in the majority of the positive results in [11, 86].

In contrast to [11, 86], we do not impose such assumptions, but instead study a model with incremental changes. When the graph is disconnected, the random walks on such graphs can have multiple stationary distributions. For technical reasons, we would always assume these vertices hold zero probabilities in their stationary distribution so that one graph can only have one stationary distribution (defined as "canonical" distribution in Section 4.3). Specifically, we consider a setting where the evolving graph model changes *randomly* and study the so-called *edge-Markovian* random graph $\mathcal{G}(n, p, q)$, which is defined as follows (see Definition 4.3.3 for a more formal description). For each edge slot, there is a two-state Markov chain that switches from off to on with probability p and from on to off with probability q. This model can be seen as a dynamic version of the Erdős-Rényi random graph, and has been studied in the context of information spreading and flooding [25, 26, 27]. While these results demonstrate that information disseminates very quickly on these dynamic graphs, analysing the convergence properties of a random walk seems to require new techniques. In the previous results, the analyses mainly applied inductive arguments, i.e., given the state of the current set of informed nodes, they derived the set for the next time step. However, when analysing the mixing time, the degree fluctuations make the use of any such arguments very difficult – from one step to another, the distribution of the walk could become worse, whereas in previous works the set of informed (or reachable) nodes can never decrease.

Furthermore, we will investigate the new mixing time of a random walk on such evolving graphs. It turns out that, as our results demonstrate, the mixing time depends crucially on the density of the graph as well as on the speed by which the graph changes.

We remark that deriving bounds on the mixing time on $\mathcal{G}(n, p, q)$ brings some unique challenges, which are not presented in the positive results of [11, 86]. The main difficulty is that there is no time-invariant stationary distribution in $\mathcal{G}(n, p, q)$ due to the changing degrees of the vertices, and the traditional notion of the mixing time must be adapted to our dynamic setting. Informally, what we ask, then, is how many steps the walk needs to take before the distance to a *time-dependent* stationary distribution becomes *small enough*. Furthermore, in contrast to static graphs, where the distance between the distribution of the walk and the stationary distribution can only decrease, in dynamic graphs the distance to the time-dependent stationary distribution might increase with time. For this reason, we also ask that the distribution of the walk remains close to a time-dependent stationary distribution for a *long enough* interval of time (for a precise definition of our notion of mixing time, see Definition 4.3.5). We believe this requirement is necessary for our definition of mixing time to be useful in potential applications.

4.2 Related works

A random walk is a basic stochastic process on graphs and a key primitive in the design of distributed algorithms [84] ranging from load balancing [52], searching [29], pagerank [24, 82, 38], voting [30, 31] and information propagation [53, 21]. Random walks are lightweight, for they do not need to maintain many complicated states. Hence they are very useful for self-organising networks. As discussed before, previous research has focused on static graphs, while there has been a growing trend for research on dynamic graphs.

Recently, [56] analysed the cover time of so-called "Edge-Uniform Stochastically-Evolving Graphs", that include our model as a special case (i.e., the history is $k = 1$). Their focus is on a process called "Random Walk with a Delay", where at each step, the walk picks a (possible) neighbour and then waits until the edge becomes present. In [56, Theorem 4], the authors also relate this process to the standard random walk, and prove a worst-case upper bound on the cover time. However, one of the key difference to [56] is that we will study the *mixing time* instead of the *cover time* because it is more complicated and also after we get the mixing time we can quickly derive some bounds on the cover time. In this line of work, [11, 86] definitely contributed significantly. Their results have been introduced in the previous section.

Other than the previously mentioned results which studied the edge-Markovian model [25, 26, 27], in [89], the authors analysed a continuous-time version of the edge-Markovian random graph. However, unlike the standard random walk, they consider a slightly different process: when the random walk tries to make a transition from a vertex u, it picks one of the $n-1$ other vertices and moves there *only if* the edge is present; otherwise it remains still. For this process, they were able to derive very tight bounds on the mixing time that establish the so-called cutoff phenomena. The same random walk was also analysed on a

dynamic graph model of the d-dimensional grid in [74, 73], and more generally, in [47].

4.3 Preliminaries

4.3.1 Random walk and conductance

In this section, we introduce relevant notations and basic results about Markov chains that we will use throughout the thesis. More background on Markov chains and random walks can be found in [58].

Let $\mathscr{G} = (G_t)_{t\in\mathbb{N}}$ be a sequence of undirected and unweighted graphs defined on the same vertex set V, with $|V| = n$, but with potentially different edge-sets E_t ($t \in \mathbb{N}$). We study (lazy) random walks on $\mathscr{G}$: suppose that at a time $t \geq 0$ a particle occupies a vertex $u \in V$. At step $t+1$, the particle will remain at the same vertex u with probability $1/2$, or will move to a random neighbour of u in G_t. In other words, it will perform a single random walk step according to a transition matrix $\mathbf{P}_t$, which is the transition matrix of a lazy random walk on G_t: $\mathbf{P}_t(u,u) = 1/2$, $\mathbf{P}_t(u,v) = 1/(2\deg_t(u))$ (where $\deg_t(u)$ is the degree of u in G_t) if there is an edge between u and v in G_t , or $\mathbf{P}_t(u,v) = 0$ otherwise.

Given an initial probability distribution $\mu_0 \colon V \to [0,1]$, which is the distribution of the initial position of the walk, the t-step distribution of a random walk on $\mathscr{G}$ is equal to $\mu_t = \mu_0\mathbf{P}_1 \cdot \mathbf{P}_2 \cdot \ldots \cdot \mathbf{P}_t$. In particular, we use μ_t^x to denote the t-step distribution of the random walk starting at a vertex $x \in V$. Hence $\mu_0^x(x) = 1$ and $\mu_0^x(y) = 0$ for $x \neq y \in V$. Furthermore, we use π_t to denote the ("canonical") probability distribution with entries equal to $\pi_t(x) = \deg_t(x)/(2|E_t|)$ for any $x \in V$. This distribution is stationary for $\mathbf{P}_t$ (i.e, it satisfies $\pi_t\mathbf{P}_t = \pi_t$) and, if G_t is connected, it is the unique stationary distribution of $\mathbf{P}_t$. If G_t is disconnected, $\mathbf{P}_t$ will have multiple stationary distribution. However, unless stated otherwise, we will consider only the "canonical" stationary distribution π_t defined above. Finally, while any individual $\mathbf{P}_t$ is *time-reversible* (it satisfies $\pi_t(x)\mathbf{P}_t(x,y) = \pi_t(y)\mathbf{P}_t(y,x)$ for any $x, y \in V$), a random walk on $\mathcal{G}$ may not. [2]

If $\mathbf{P}$ is a transition matrix of a reversible Markov chain, it has n real eigenvalues, which we denote with $-1 \leq \lambda_n(\mathbf{P}) \leq \cdots \leq \lambda_1(\mathbf{P}) = 1$. If $\mathbf{P}$ is the transition matrix of a lazy random walk on a graph G, it holds that $\lambda_n(\mathbf{P}) \geq 0$. Moreover, $\lambda_2(\mathbf{P}) < 1$ if and only if G is connected.

For two probability distributions $f, g \colon V \to [0,1]$, the *total variation distance* between f and g is defined as $\|f-g\|_{TV} := \frac{1}{2}\sum_{x\in V}|f(x)-g(x)|$. We denote with $\|f\|_2 = \left(\sum_{x\in V} f^2(x)\right)^{1/2}$ and $\|f\|_\infty = \max_{x\in V}|f(x)|$ the standard ℓ_2 and ℓ_∞ norms of f. Given a

[2] For example, it might happen that $\mathbf{P}_1 \cdots \mathbf{P}_t(x,y) > 0$ while $\mathbf{P}_1 \cdots \mathbf{P}_t(y,x) = 0$. This cannot happen in the "static" case where $\mathbf{P}_1 = \cdots = \mathbf{P}_t = \mathbf{P}$ with $\mathbf{P}$ reversible.

probability distribution $\pi: V \to \mathbb{R}_+$, we also define the $\ell_2(\pi)$-norm as

$$\|f\|_{2,\pi} := \sqrt{\sum_{x \in V} f^2(x)\pi(x)}$$

By Jensen's inequality, it holds for any f, g that $\|f - g\|_{TV} \leq \|f - g\|_{2,\pi}$. The lemma below relates the decrease in the distance to stationarity after one random walk step to the spectral properties of its transition matrix.

Lemma 4.3.1 (Lemma 1.13 in [65], rephrased)**.** *Let* $\mathbf{P}$ *be the transition matrix of a lazy random walk on a graph* $G = (V, E)$ *with stationary distribution* π*. Then, for any* $f : V \to \mathbb{R}$*, we have that*

$$\left\| \frac{f\mathbf{P}}{\pi} - \mathbf{1} \right\|_{2,\pi}^2 \leq \lambda_2(\mathbf{P})^2 \left\| \frac{f}{\pi} - \mathbf{1} \right\|_{2,\pi}^2 .$$

In the lemma above and throughout the thesis, a division between two functions is to be understood entry-wise, while $\mathbf{1}$ refers to a function always equal to one. An important quantity which can be used to obtain bounds on $\lambda_2(\mathbf{P})$ is the *conductance* of G, which is defined as follows.

Finally, we use the notation $o_n(1)$ to denote any function $f : \mathbb{N} \to \mathbb{R}$ such that $\lim_{n \to +\infty} f(n) = 0$. We often drop the subscript n.

4.3.2 Dynamic graph models

In this section we formally introduce the random models of (dynamic) graphs that are the focus of this work. We start by recalling the definition of the Erdős-Rényi model of random (static) graphs.

Definition 4.3.2 (Erdős-Rényi model)**.** $G = (V, E) \sim \mathcal{G}(n, p)$ *is a random graph such that* $|V| = n$ *and the* $\binom{n}{2}$ *possible edges appear independently, each with probability* p*.*

We now introduce the *edge-Markovian* model of dynamic random graphs, which has been studied both in the context of information spreading in networks [25, 26] and random walks [56]. This model is the focus of our work.

Definition 4.3.3 (edge-Markovian model)**.** *Given a starting graph* G_0*, we denote with* $(G_t)_{t \in \mathbb{N}} \sim \mathcal{G}(n, p, q)$ *a sequence of graphs such that* $G_t = (V, E_t)$*, where* $V = \{1, \ldots, n\}$ *and, for each* $t \in \mathbb{N}$*, any pair of distinct vertices* $u, v \in V$ *will be connected by an edge in* G_t *independently at random with the following probability:*

$$\mathbb{P}\left[\{u, v\} \in E_{t+1} \mid G_t\right] = \begin{cases} 1 - q & \text{if } \{u, v\} \in E_t \\ p & \text{if } \{u, v\} \notin E_t. \end{cases}$$

Notice that different choices of a starting graph G_0 will induce different probability distributions over $(G_t)_{t\in\mathbb{N}}$. In general, we try to study $\mathcal{G}(n,p,q)$ by making the fewest possible assumptions on our choice of G_0. Moreover, as pointed out for example in [56], the probability that we sample a graph $(G_t)_{t\in\mathbb{N}} \sim \mathcal{G}(n,p,q)$ is the same as sampling a graph from $\mathcal{G}(n,\tilde{p})$ with $\tilde{p} = p/(p+q)$. We leave considerations about the speed of this convergence and how this affects our choice of G_0 to Section 4.9 and, in particular, Remark 4.9.2 .

4.3.3 Mixing time of random walks on dynamic graphs

One of the most studied quantities in the literature about time-homogeneous (i.e., static) Markov chains (random walks included) is the mixing time, i.e., the time it takes for the distribution of the chain to become close to stationarity. Formally, it is defined as follows.

Definition 4.3.4 (Mixing time for time-homogeneous Markov chains)**.** *Let μ_t^x be the t-step distribution of a Markov chain with state space V starting from $x \in V$. Let π be its stationary distribution. For any $\epsilon > 0$, the ϵ-mixing time is defined as*

$$t_{\text{mix}}(\epsilon) := \min\{t \in \mathbb{N} : \max_{x\in V} \|\mu_t^x - \pi\|_{TV} \leq \epsilon\}.$$

A basic fact in random walk theory states that a lazy random walk on a connected undirected graph $G = (V,E)$ has always a finite mixing time. In particular, if $|V| = n$, $t_{\text{mix}}(1/4) = O(n^3)$. Moreover, considering a different ϵ does not significantly change the mixing time: for any $\epsilon > 0$, $t_{\text{mix}}(\epsilon) = O(t_{\text{mix}}(1/4)\log(1/\epsilon))$ (see, e.g., [58]). Also, it is a well-known fact that $\|\mu_t^x - \pi\|_{TV}$ is non-increasing. This is also similar to the load balancing discrepancy convergence in the previous chapter.

However, in the case of random walks on dynamic graphs, convergence to a time-invariant stationary distribution does not, in general, happen. For this reason, other works have studied alternative notions of mixing for dynamic graphs, such as merging [77], which happens when a random walk "forgets" the vertex where it started. In this work, instead, we focus on a different approach that we believe best translates the classical notion of mixing from the static to the dynamic case. More precisely, let us consider a dynamic sequence of graphs $(G_t)_{t\in\mathbb{N}}$ with corresponding stationary distributions $(\pi_t)_{t\in\mathbb{N}}$. Our goal is to establish if there exists a time t such that the distribution μ_t of the walk at time t is close to π_t. Moreover, to make this notion of mixing useful in possible applications, we require that μ_s remains close to π_s for a reasonably large number of steps $s \geq t$. Formally, we introduce the following definition of mixing time for dynamic graph sequences.

Definition 4.3.5 (Mixing time for dynamic graph sequences)**.** *Let $\mathcal{G} = (G_t)_{t\in\mathbb{N}}$ be a dynamic graph sequence on a vertex set V, $|V| = n$. The mixing time of a random walk*

in $\mathcal{G}$ is defined as

$$t_{\text{mix}}(\mathcal{G}) = \min\left\{t \in \mathbb{N} : \forall s \in [t, t+\sqrt{n}],\ \forall x \in V,\ \ \|\mu_s^x - \pi_s\|_{TV} = o_n(1)\right\},$$

where π_s is the stationary distribution of a random walk in G_s, and μ_s^x is the s-step distribution of a random walk in $\mathcal{G}$ that starts from $x \in V$.

First observe we require that the total variation distance between μ_s and π_s goes to zero as the number of vertices increases.[3] This is motivated by the fact that the distance to stationarity, unlike in the static case, might never drop beyond a certain threshold: for this reason, we explicitly require that such threshold becomes smaller and smaller with an increasing number of vertices. Secondly, we require that such distance remains small for $\sqrt{n}$ steps (recall n is the number of vertices in the graph). This is due to the fact that, for all dynamic graph models we consider, we cannot hope for such distance to stay small arbitrarily long. However, we believe that $\sqrt{n}$ steps is a long enough period of time for mixing properties to be useful in applications.

Since our goal is to study the mixing property of $\mathcal{G}(n,p,q)$, we now introduce a definition of mixing time for edge-Markovian models that takes into account the *probabilistic nature* of such graph sequences. Essentially, we say that the mixing time of $\mathcal{G}(n,p,q)$ is t if a random walk on a dynamic sequence of graphs sampled from $\mathcal{G}(n,p,q)$ mixes (according to the previous definition) in t steps *with high probability* over the sampled dynamic graph sequence.

Definition 4.3.6 (Mixing time for edge-Markovian models)**.** *Given an edge-Markovian model $\mathcal{G}(n,p,q)$, its mixing time is defined as*

$$t_{\text{mix}}(\mathcal{G}(n,p,q)) = \min\left\{t \in \mathbb{N}\colon\ \mathbb{P}_{\mathcal{G}\sim\mathcal{G}(n,p,q)}\left[t_{\text{mix}}(\mathcal{G}) \le t\right] \ge 1 - o_n(1)\right\}.$$

The reason of having this probabilistic version is that a "bad" graph (e.g., a sparse and disconnected graph) can appear in the sequence with probability 1 given sufficiently long time, which can cause a huge gap between the walk distribution and the stationary distribution. Also it is always possible to have a sequence of poorly connected graphs. Hence we use the definition above to emphasise that in practice it is enough to know the probability that the sampled graph sequence has a quick mixing time is large.

Finally, we remark that, while in static graphs connectivity is a necessary prerequisite to mixing, random walks on sequences of disconnected dynamic graphs might nonetheless exhibit mixing properties. Examples of this behavior were studied in [86].

[3] We are implicitly assuming there is an infinite family of dynamic graph sequences with increasing n.

4.4 Main results

We study the mixing properties of random walks on edge-Markovian random graphs $\mathcal{G}(n, p, q)$. In particular, we consider six different settings of parameters p and q, which separates edge-Markovian models based on how fast graphs change over time (slowly vs. fast changing), and how dense graphs in the dynamic sequence are (sparse vs. semi-sparse vs dense).

As noted in previous works (see, e.g., [25]), a dynamic sequence sampled from $\mathcal{G}(n, p, q)$ is equivalent to a sample in the Erdős-Rényi random graph $\mathcal{G}(n, \tilde{p})$ where $\tilde{p} = \frac{p}{p+q}$ (for the sake of completeness, we give a proof in Section 4.9). We use the expected degree in such a random graph, which is equal to $d = (n-1)\tilde{p}$, to separate edge-Markovian models according to their density as follows:

1. **Sparse** $d = o(\log n)$
2. **Semi-sparse** $d = \Theta(\log n)$
3. **Dense** $d = \omega(\log n)$.

Notice that the sparse regime corresponds to random graphs with density below the connectivity threshold of Erdős-Rényi random graphs.

We further separate edge-Markovian models based on how fast they change over time. Let $\delta = \binom{n}{2}\tilde{p}q + \binom{n}{2}(1-\tilde{p})p$ be the expected number of changes at each step, when starting from a stationary initial graph $G_0 \sim \mathcal{G}(n, \tilde{p})$. We consider the following two opposite regimes.

1. **Fast-changing** $\delta = \Theta(dn)$.
2. **Slowly-changing** $\delta = O(\log n)$

Notice that the fast-changing regime corresponds to graphs for which a constant fraction of edges change at each step in expectation.

For a cleaner exposition, we consider six different settings based on fast-changing vs. slowly-changing rates and dense vs. medium vs. sparse graphs (Table 4.4). The density of the graph refers to the expected degree of the graph sampled from the *stationary graph distribution* (Sec. 4.3) of the model. Moreover this quantifies the expected degree $d = n\tilde{p}$. For reasons of space, we will not analyse all combinations of p and q. Also note that some choices of p and q are not interesting, for example, small values of p (and q) may trap the walk inside a small region for a long time.

The main results of our work are presented in Table 4.4. Here, we assume G_0 is sampled from the stationary graph distribution $\mathcal{G}(n, \tilde{p})$. In the fast-changing regime, as

[4] In this regime we are not able to prove finite mixing time. However, we show that the distribution of the walk will "flatten out" after $O(\log n)$ steps. We refer to this behavior as *coarse* mixing.

	Fast-changing $\delta = \Theta(dn)$	**Slowly-changing** $\delta = O(\log n)$
Sparse $d \in [1, o(\log n)]$	$t_{\text{mix}} = \infty$ Thm 4.4.1	$t_{\text{mix}} = \Omega(n)$ Proposition 4.4.4
Semi-sparse $d = \Theta(\log n)$	Coarse mixing[4] in $O(\log n)$ Prop 4.5.2	$t_{\text{mix}} = O(\log n)$, Thm 4.4.3
Dense $d \in [\omega(\log n), n/2]$	$t_{\text{mix}} = O(\log n)$ Thm 4.4.2	

Table 4.1: Summary of our main results (informal). See referenced theorems for the precise and complete statements.

highlighted in Remark 4.9.2, this is without loss of generality. For slow-changing models, instead, different choices of G_0 can result in drastically different outcomes with regard to the mixing time. For ease of presentation, we assume in Table 4.4 that $G_0 \sim \mathcal{G}(n, \tilde{p})$, but this assumption can usually be relaxed, and we refer to the full statement of the corresponding results for our actual assumptions on G_0.

Next we formally state the four main results of our work. The formal definitions of mixing time for random walk on dynamic graphs will be presented in Section 4.3 (see in particular Definition 4.3.5 and Definition 4.3.6). The first theorem is a negative result that tells us that, for fast-changing and sparse edge-Markovian graphs, random walks don't have finite mixing time. Its proof will be presented in Section 4.5.1.

Theorem 4.4.1 (Fast-changing and sparse, no mixing)**.** *Let* $p = \Theta\left(\frac{1}{n}\right)$ *and* $q = \Omega(1)$*. Then,* $t_{\text{mix}}(\mathcal{G}(n, p, q)) = \infty$*.*

The following theorem is, instead, a positive result that establishes fast mixing time in the dense and fast-changing regime. Its proof is presented in Section 4.5.2.

Theorem 4.4.2 (Fast-changing and dense, fast mixing)**.** *Let* $p = \omega(\log n/n)$ *and* $q = \Omega(1)$*. Then,* $t_{\text{mix}}(\mathcal{G}(n, p, q)) = O(\log n)$*.*

The only case missing in the fast-changing regime is the semi-sparse case, where nodes have average degree $d = \Theta(\log n)$. We do not have a definitive answer on the mixing time of random walks in such case, however, we do have a partial result that guarantees at least that random walk distributions will be "well spread" over a large support after $O(\log n)$ steps (we call this behavior *coarse mixing*). This statement can be made formal by considering the ℓ_2-norm of the distribution of the walk. Because of its technical nature, we defer the formal statement to Proposition 4.5.2.

We now turn our attention to the slowly-changing regime, where at most $\delta = O(\log n)$ edges are created and removed at each step. Unlike the results for the fast-changing regime, where the choice of the starting graph G_0 does not really affect the mixing time of

a random walk (see Section 4.9 and Remark 4.9.2 for a discussion), in the slowly-changing regime the choice of G_0 will affect the properties of G_t for a large number of steps t.

The following theorem shows that in the slowly-changing and dense regime, under mild conditions on the starting graph $G_0 = (V, E_0)$ (which are satisfied for G_0 drawn from the limiting distribution of dense $\mathcal{G}(n, p, q)$), random walks will mix relatively fast. We use $E_0(S, V \setminus S)$ to indicate the set of edges in G_0 between a subset of vertices $S \subset V$ and its complement, and $\Phi_{G_0}$to indicate the minimum conductance of G_0 (see Definition 2.4.1).

Theorem 4.4.3 (Slowly-changing and dense, fast mixing)**.** *Let $d = \Omega(\log n)$, $p = O(\log n/n^2)$, and $q = O(\log n/(dn))$. Let the following assumptions on the starting graph $G_0 = (V, E_0)$ be satisfied for large enough constants $c_1, c_2, c_3 > 0$.*

1. $\deg_0(x) = \Theta(d)$ *for any* $x \in V$*;*
2. $|E_0(S, V \setminus S)| \geq c_2 \log n|S|$*, for any* $S \subset V$ *with* $|S| \leq c_1 \log n$*;*
3. $\Phi_{G_0} \geq c_3 \log d/d$.

Then, $t_{\text{mix}}(\mathcal{G}(n, p, q)) = O(\log n/\Phi^2_{G_0})$.

Let us briefly discuss the assumptions and results of Theorem 4.4.3.First of all notice that the parameters p and q are defined so that the average degree $d = \Omega(\log(n))$ and the number of changes in the graph at each step is $\delta = O(\log(n))$. Assumption (1) just require the degree of the vertices in G_0 to be of the same order as the degree of the vertices in the limiting graph $\mathcal{G}(n, \tilde{p})$. Assumption (2) guarantees that for any small set S there are enough edges going from S to the rest of the graph. Assumption (3) is a mild condition on the conductance of G_0. These two conditions ensure that the conductance of G_t will not be much lower than the conductance of G_0 for a large number of steps t. Finally, notice that $O(\log n/\Phi^2_{G_0})$ is a classic bound for the mixing time of a *static* random walk on G_0. Theorem 4.4.3 essentially states that, if the three assumptions are satisfied, the mixing time of a random walk on $\mathcal{G}(n, p, q)$ will not be much larger. In particular, all the three assumptions are satisfied for a starting graph $G_0 \sim \mathcal{G}(n, \tilde{p})$ with $\tilde{p} = p/(p + q)$. Furthermore, in such case $t_{\text{mix}}(\mathcal{G}(n, p, q)) = O(\log n)$. The proof of this theorem can be found in Section 4.6.1.

We conclude this section by stating our result in the slowly-changing and dense regime. We prove a negative result: we show that the mixing time of $\mathcal{G}(n, p, q)$ is at least linear in n.

Proposition 4.4.4 (Slowly-changing and sparse, slow mixing)**.** *Let $p = O(1/n^2)$ and $q = \omega(1/(n \log n))$. Consider a random walk on $\mathcal{G}(n, p, q)$ with starting graph $G_0 \sim \mathcal{G}(n, \tilde{p})$ with $\tilde{p} = p/(p + q)$. Then, $t_{\text{mix}}(\mathcal{G}(n, p, q)) = \Omega(n)$.*

4.5 Results for the fast-changing case

4.5.1 Negative result in the sparse and fast-changing case

In this section we consider random walks on sparse and fast-changing edge-Markovian graphs. In particular, we study $\mathcal{G}(n,p,q)$ with $0 < q = \Omega(1)$ and $p = \Theta(\frac{1}{n})$. Since $\Omega(1)$, by Remark 4.9.2, we can restrict ourselves to consider the case where $G_0 \sim \mathcal{G}(n,\tilde{p})$ with $\tilde{p} = p/(p+q)$. We will show a negative result on the mixing of random walks in this regime: no matter how small is the distance between μ_t and π_t, the total variation distance will increase to a positive constant with constant nonzero probability.

Theorem 4.4.1 (Fast-changing and sparse, no mixing)**.** *Let $p = \Theta\left(\frac{1}{n}\right)$ and $q = \Omega(1)$. Then, $t_{\text{mix}}(\mathcal{G}(n,p,q)) = \infty$.*

The key idea behind this result is that due to the fast-changing nature, the degrees of nodes change rapidly. In particular, for a linear number of nodes u, there is at least one neighbour $v_{\min} \in \Gamma_t(u)$ whose degree may change from one constant in round t to, basically any other constant (this also makes use of the assumption on p, ensuring that the graph is sparse). The proof then exploits that, due to the "unpredictable" nature of this change, the probability mass received by $v_{\min}$ in round $t+1$ is likely to cause a significant difference between $\mu_{t+1}(u)$ and $\pi_{t+1}(u)$. Since this holds for a linear number of nodes u, we obtain a sufficiently large lower bound on the total variation distance, and the theorem is established.

Proof of Theorem 4.4.1. We will prove that no matter how small the distance between μ_t and π_t is, the total variation distance between μ_{t+1} and π_{t+1} will increase to a positive constant with constant nonzero probability. This will yield the theorem.

In this regime, at a time t the graph has converged to $\mathcal{G}(n,\tilde{p})$, and when the graph has isolated nodes, the stationary distribution of the random walk on a disconnected graph is not unique. Recall that in this case, we choose $\pi_t(u) = \deg(u)/2|E_t|$ as the stationary distribution.

For any node u, we have

$$\begin{aligned}|\mu_{t+1}(u) - \pi_{t+1}(u)| &= \left| \left(\sum_{v \in \Gamma_{t+1}(u)} \pi_t(v) \cdot \frac{1}{\deg_{t+1}(v)} \right) - \frac{\deg_{t+1}(u)}{2|E_{t+1}|} \right| \\ &= \left| \sum_{v \in \Gamma_{t+1}(u)} \left(\frac{\deg_t(v)}{2|E_t|} \frac{1}{\deg_{t+1}(v)} - \frac{1}{2|E_{t+1}|} \right) \right|,\end{aligned}$$

where by Chernoff bound, we know $|E_{t+1}|/|E_t|$ is close to 1 and both of them are $\Theta(n)$ in this regime. Hence we can use $|E| = \binom{n}{2}\tilde{p}$ to replace either of them to get an approximation. So the distance is contributed by $\deg_t(v)/\deg_{t+1}(v)$. In this proof, we assume $p = \frac{1}{n}$ and

$q = \frac{1}{2}$. The same result still holds for $p = c/n$ with small constant c. If we increase c, some constants and quantities below should be changed according to how large c is. If it gets too large, we may consider that it will be classified into the dense case instead of this case.

For a vertex $u \in V$, let $v_{\min} \in \Gamma_t(u)$ be the neighbour of u with the smallest degree in G_t, and assume u is a vertex so that $\deg_t(v_{\min}) \leq 20$. Let us define $\alpha := \sum_{v \in \Gamma_{t+1}(u) \setminus \{v_{\min}\}} \left(\frac{\deg_t(v)}{\deg_{t+1}(v)} - 1 \right)$, the contribution by all neighbours of u instead of $v_{\min}$. Also let us define $\beta := \sum_{v \in \Gamma_{t+1}(u)} \mathbf{1}_{\{\{v, v_{\min}\} \in E_{t+1}\}}$, the number of neighbours of $v_{\min}$ in $\Gamma_{t+1}(u)$ and $\gamma := \sum_{v \in V \setminus \Gamma_{t+1}(u)} \mathbf{1}_{\{\{v, v_{\min}\} \in E_{t+1}\}}$, the number of neighbours of $v_{\min}$ in $V \setminus \Gamma_{t+1}(u)$. Let us define the following events:

$$
\begin{aligned}
\mathcal{A} &:= \{|E_{t+1}| = (1 \pm \sqrt{10 \log n / n})|E|\} \cap \{|E_t| = (1 \pm \sqrt{10 \log n / n})|E|\}, \\
\mathcal{B} &:= \{\beta \leq 80\}, \\
\mathcal{C} &:= \left\{ \left| \alpha + \left(\frac{\deg_t(v_{\min})}{\beta + \gamma} - 1 \right) \right| \geq c_4 \right\}, \\
\mathcal{D} &:= \{v_{\min} \in \Gamma_{t+1}(u)\},
\end{aligned}
$$

where $|E| = \binom{n}{2}\tilde{p}$ is the expected number of edges of a stationary graph, and $c_4 > 0$ is a proper constant that is defined later.

Using Chernoff bound, we have that

$$
\mathbb{P}[\neg \mathcal{A}] \leq \exp\left(\frac{10 \log n |E|}{2n} \right) = O(n^{-c}),
$$

for some constant c.

Note that $\mathbb{P}[\mathcal{A}] \geq 1 - o(1)$, $\mathbb{P}[\mathcal{B}] \geq 3/4$ and $\mathbb{P}[\mathcal{D}] = 1 - q > 0$ which is also a constant probability in this regime. If the event $\mathcal{A}$ holds, then

$$
\begin{aligned}
&|\mu_{t+1}(u) - \pi_{t+1}(u)| \\
&= \sum_{v \in \Gamma_{t+1}(u)} \frac{1}{2|E_t|} \left(\frac{\deg_t(v)}{\deg_{t+1}(v)} - 1 \right) \pm O(\log^2 n / \sqrt{n}).
\end{aligned}
$$

Further, assuming that $\mathcal{D}$ holds, we can lower bound

$$
\sum_{v \in \Gamma_{t+1}(u) \setminus \{v_{\min}\}} \frac{1}{2|E_t|} \left(\frac{\deg_t(v)}{\deg_{t+1}(v)} - 1 \right) + \frac{1}{2|E_t|} \left(\frac{\deg_t(v_{\min})}{\deg_{t+1}(v_{\min})} - 1 \right).
$$

Now let us expose all degrees for $v \in \Gamma_{t+1}(u) \setminus \{v_{\min}\}$, subject to event $\mathcal{B}$ holding. Hence the above formula is equal to

$$
\frac{1}{2|E_t|} \cdot \left(\alpha + \frac{\deg_t(v_{\min})}{\beta + \gamma} - 1 \right).
$$

Recall that α and $\beta \leq 80$ are arbitrary. Further, γ is independent of α and β. Also for any constant $c_1 > 0$, $\mathbb{P}[\gamma = c_1] > c_2$ for some other constant $c_2 > 0$. Note in order for $|\mu_{t+1}(u) - \pi_{t+1}(u)|$ not to contribute significantly to the total variation distance, we must have

$$\alpha + \frac{\deg_t(v_{\min})}{\beta + \gamma} - 1 = 0,$$

which is equivalent to

$$\gamma = \frac{\deg_t(v_{\min})}{1 - \alpha} - \beta.$$

There is at most one possible (positive integer) value for γ that solves this equation after α, β have been revealed. In particular, there is at least one constant $c_3 > 0$ so that

$$\left|\alpha + \frac{\deg_t(v_{\min})}{\beta + c_3} - 1\right| > c_4,$$

for some other constant $c_4 > 0$ because $v_{\min} \leq 20$ and $\beta \leq 80$. This proves $\mathbb{P}[\mathcal{C} \mid \mathcal{B}] \geq c_5$. Further,

$$\begin{aligned}
\mathbb{P}[\mathcal{A} \cap \mathcal{C}] &\geq \mathbb{P}[\mathcal{A}] - \mathbb{P}[\neg\mathcal{C}] \\
&\geq \mathbb{P}[\mathcal{A}] - 1 + \mathbb{P}[\mathcal{C}] \\
&\geq \mathbb{P}[\mathcal{A}] - 1 + \mathbb{P}[\mathcal{C} \wedge \mathcal{B}] \\
&\geq \mathbb{P}[\mathcal{A}] - 1 + \mathbb{P}[\mathcal{B}] \cdot \mathbb{P}[\mathcal{C} \mid \mathcal{B}] \\
&\geq -o(1) + \frac{3}{4} \cdot c_5 > 0.
\end{aligned}$$

Hence with constant probability, any vertex $u \in V$ in G_t that has a neighbour $v_{\min}$ with $\deg_t(v_{\min}) \leq 20$ will contribute $\Omega(1/|E|)$ to the ℓ_1-norm with constant probability $c_6 > 0$. Let us now lower bound the number of such vertices,

$$S_t := \left\{ u \in V(G_t) \colon \min_{v \in \Gamma_t(u)} \{\deg(v)\} \leq 20 \right\}.$$

Fix any vertex $u \in V$. With constant probability $C_1 > 0$, it has at least one neighbour, say, w. Further, that neighbour w will have at most 19 neighbours (besides u) with constant probability $C_2 > 0$. Hence $\mathbb{E}[|S_t|] \geq (C_1 C_2) \cdot n$. Using Markov's inequality, it follows that

$$\mathbb{P}[n - |S_t| \geq (1 + C_1C_2/2) \cdot (1 - C_1C_2) \cdot n] \leq \frac{1}{1 + C_1C_2/2}.$$

Rearranging the above, it follows that with constant probability $C_3 := \frac{1}{1+C_1C_2/2} > 0$, we have $|S_t| \geq (C_1C_2/2) \cdot n$. Hence the overall contribution of S_t is at least $\Omega(n) \cdot \Omega(1/|E|) = \Omega(1)$ (as $|E| = \Theta(n)$ in this regime), with some constant probability $c_7 > 0$. $\square$

4.5.2 Positive results in the dense and fast-changing case

In this section we analyse the mixing properties of $\mathcal{G}(n,p,q)$ for $p = \Omega(\log n/n)$ and $q = \Omega(1)$. Since q is large, for simplicity we will assume throughout this section that $G_0 \sim \mathcal{G}(n,\tilde{p})$, where $\tilde{p} = \frac{p}{p+q}$ (see Remark 4.9.2 for an explanation of why this is not a restriction). The main result of this section is the following theorem.

Theorem 4.4.2 (Fast-changing and dense, fast mixing)**.** *Let $p = \omega(\log n/n)$ and $q = \Omega(1)$. Then, $t_{\mathrm{mix}}(\mathcal{G}(n,p,q)) = O(\log n)$.*

While in this thesis we study for simplicity only lazy random walks on graphs, to prove Theorem 4.4.2, however, we need to introduce *simple* random walks on graphs: given a graph $G = (V,E)$, a simple random walk on G has transition matrix Q such that, for any $x,y \in V$, $Q(x,y) = 1/\deg(x)$ if $\{x,y\} \in E$, $Q(x,y) = 0$ otherwise. The following lemma, whose proof is the main technical part of the section, shows that if the *simple* random walk on a sequence of graphs $\mathcal{G} = (G_t)_{t\in\mathbb{N}}$ exhibits strong expansion properties, and the time-varying stationary distribution is always close to uniform, then a *lazy* random walk on $\mathcal{G}$ will be close to the stationary distribution of G_t for any t large enough. We remark that a strong expansion condition on lazy random walks can never be satisfied; luckily, we just need this strong expansion condition to hold for their simple counterpart.

Lemma 4.5.1. *Let $(G_t)_{t\in\mathbb{N}}$ be a sequence of graphs, and $(P_t)_{t\in\mathbb{N}}$ (resp. $(Q_t)_{t\in\mathbb{N}}$) the corresponding sequence of transition matrices for a* lazy *(resp.* non-lazy*) random walk. Assume there exists $1 < C = O(1)$ such that, for any $t \geq 1$ and any $x \in V$, $1/(C\cdot n) \leq \pi_t(x) \leq C/n$. Moreover, also assume that, for any $t \in \mathbb{N}$, $\max\{|\lambda_2(Q_t)|, |\lambda_n(Q_t)|\} \leq \lambda = o(1)$. Then, there exists an absolute constant C' such that, w.h.p., for any $t \geq C'\log n$ and any starting distribution μ_0,*

$$\left\| \frac{\mu_t}{\pi_t} - \mathbf{1} \right\|^2_{2,\pi_t} \leq 10C^2(C-1)^2,$$

where $\mu_t = \mu_0 P_1 \cdots P_t$.

Proof of Lemma 4.5.1. We first relate the $\ell_2(\pi)$ distance to stationarity to the ℓ_2 distance from the uniform distribution. We start by observing that, for any $t \in \mathbb{N}$, by our assumptions on π_t, it holds that

$$\begin{aligned}
\|\pi_t - \mathbf{1}/\mathbf{n}\|_2^2 &= \sum_{x\in V} \left(\pi_t(x) - \frac{1}{n}\right)^2 \\
&\leq \max\left\{\left(1-\frac{1}{C}\right)^2, (C-1)^2\right\} \cdot \frac{n}{n^2} \\
&= \frac{(C-1)^2}{n}.
\end{aligned}$$

Then, for any probability distribution p, we have that

$$\begin{aligned}\left\|\frac{p}{\pi_t}-\mathbf{1}\right\|_{2,\pi_t}^2 &= \sum_x \frac{(p(x)-\pi_t(x))^2}{\pi_t(x)}\\ &\leq Cn\cdot\|p-\pi_t\|_2^2\\ &\leq 2Cn\cdot\left(\|p-\mathbf{1}/\mathbf{n}\|_2^2+\|\pi_t-\mathbf{1}/\mathbf{n}\|_2^2\right) \qquad (4.1)\\ &\leq 2Cn\cdot\left(\|p-\mathbf{1}/\mathbf{n}\|_2^2+(C-1)^2/n\right)\\ &= 2Cn\cdot\|p-\mathbf{1}/\mathbf{n}\|_2^2+2C(C-1)^2, \qquad (4.2)\end{aligned}$$

where (4.1) holds by the triangle inequality and the basic inequality $(a+b)^2\leq 2a^2+2b^2$. Analogously, it holds that

$$\begin{aligned}\left\|\frac{p}{\pi_t}-\mathbf{1}\right\|_{2,\pi_t}^2 &\geq (n/C)\cdot\left(\frac{1}{2}\|p-\mathbf{1}/\mathbf{n}\|_2^2-3\|\pi_t-\mathbf{1}/\mathbf{n}\|_2^2\right) \qquad (4.3)\\ &\geq (n/C)\cdot\left(\frac{1}{2}\|p-\mathbf{1}/\mathbf{n}\|_2^2-\frac{3}{n}(C-1)^2\right)\\ &\geq (n/C)\cdot\|p-\mathbf{1}/\mathbf{n}\|_2^2-3(C-1), \qquad (4.4)\end{aligned}$$

where (4.3) holds by the triangle inequality and the basic inequality $(a-b)^2\geq a^2/2-3b^2$.

Notice that the distance to the uniform distribution does not change if at step t we perform a lazy step, which happens with probability $1/2$. Conditioning on the fact that we don't take a lazy step, at time t we can bound the decrease in the distance to the uniform distribution as follows:

$$\begin{aligned}\left\|\frac{\mu_{t+1}}{\pi_{t+1}}-\mathbf{1}\right\|_{2,\pi_{t+1}}^2 &\leq \lambda\cdot\left\|\frac{\mu_t}{\pi_{t+1}}-\mathbf{1}\right\|_{2,\pi_{t+1}}^2\\ &\leq \lambda\cdot\left(2Cn\cdot\|\mu_t-\mathbf{1}/\mathbf{n}\|_2^2+2C(C-1)^2\right), \qquad (4.5)\end{aligned}$$

where the first inequality follows from Lemma 4.3.1 and the second by (4.2). Moreover, we have that

$$\begin{aligned}n\cdot\|\mu_{t+1}-\mathbf{1}/\mathbf{n}\|_2^2 &\leq C\left\|\frac{\mu_{t+1}}{\pi_{t+1}}-\mathbf{1}\right\|_{2,\pi_{t+1}}^2+3C(C-1)\\ &\leq o(\|\mu_t-\mathbf{1}/\mathbf{n}\|_2^2+C^2(C-1)^2)+3C(C-1), \qquad (4.6)\end{aligned}$$

where the first inequality follows from (4.4), and the last from (4.5) and the assumption $\lambda=o(1)$.

This implies that whenever $n\cdot\|\mu_t-\mathbf{1}/\mathbf{n}\|_2^2$ is large enough (e.g., $n\cdot\|\mu_t-\mathbf{1}/\mathbf{n}\|_2^2\geq 4C(C-1)$), if we condition on the walk not taking a lazy step at time t, the distance to the uniform distribution will shrink significantly (this follows because $\lambda=o(1)$). Therefore,

we just need $O(\log n)$ non-lazy steps for such distance to become small. Hence, after $t = O(\log n)$ steps, it holds w.h.p. that

$$n \cdot \|\mu_t - \mathbf{1}/\mathbf{n}\|_2^2 \leq 4C(C-1), \tag{4.7}$$

which also implies by (4.2) that

$$\left\| \frac{\mu_t}{\pi_t} - \mathbf{1} \right\|_{2,\pi_t}^2 \leq 10C^2(C-1)^2.$$

Moreover, after $O(\log(n))$ steps, this distance will continue to be small. In fact, let μ_t satisfy (4.7). If we condition on taking a lazy step at time $t+1$, such distance will not change. If instead we take a non-lazy step, by (4.6),

$$n \cdot \|\mu_t - \mathbf{1}/\mathbf{n}\|_2^2 \leq 3C(C-1) + o(C^2(C-1)^2) \leq 4C(C-1),$$

and, therefore, the distance to the uniform distribution again satisfies (4.7). The lemma follows by applying (4.2) once again. □

We now show how it can be used to derive Theorem 4.4.2. First recall that since we are assuming $G_0 \sim \mathcal{G}(n, \tilde{p})$, all graphs in the sequence $(G_t)_{t\in\mathbb{N}}$ are sampled (non-independently) from $\mathcal{G}(n, \tilde{p})$ (see Section 4.9). Furthermore, for any $t \in \mathbb{N}$, the assumptions of Theorem 4.4.2 on $\lambda_2(Q_t)$ and $\lambda_n(Q_t)$ are satisfied with probability $1 - o(1/n^2)$ for any graph sampled from $\mathcal{G}(n, \tilde{p})$ with $\tilde{p} > 2\log n/n$ by [49, Theorem 1.1]. Moreover, for $\tilde{p} = \omega(\log n/n)$, by standard Chernoff bounds argument we can show that, with probability $1-o(1/n^2)$, all vertices of a graph sampled from $\mathcal{G}(n, \tilde{p})$ have degree $(1+o_n(1))n\tilde{p}$. This implies that, for any t, w.h.p, the stationary distribution of G_t satisfies the assumptions of Lemma 4.5.1 with $C = 1 + o(1)$, which yields Theorem 4.4.2.

It is natural to ask if we can relax the condition on p. Assume for example that p, q are such that $\tilde{p} = p/(p+q) > 2\log n$. By [49, Theorem 1.1], the conditions on λ are still satisfied. However, it only holds that $C = \Theta(1)$. Therefore, Lemma 4.5.1 can only establish that the $\ell_2(\pi_t)$-distance to stationarity is a constant (potentially larger than 1). This, unfortunately, does not give us any meaningful bound on the total variation distance. However, if the ℓ_2-distance between two distributions μ and π is small, $\mu(x)$ cannot be much larger than $\pi(x)$. In a sense, this result can be interpreted as a *coarse* mixing property. This is summarised in the following proposition.

Proposition 4.5.2. *Let $(G_t)_{t\in\mathbb{N}} \sim \mathcal{G}(n, p, q)$ with $p/(p+q) > 2\log n/n$ and $q = \Omega(1)$. Let π_t be the stationary distribution of G_t. Then, there exists absolute constants $c_1, c_2 > 0$ such that, for any starting distribution μ_0 and any $c_1 \log n \leq t \leq \sqrt{n} + c_1 \log n$, it holds*

that

$$\mathbb{P}\left[\left\|\frac{\mu_t}{\pi_t}-\mathbf{1}\right\|_{2,\pi_t}^2 \leq c_2\right] \geq 1-o_n(1).$$

4.6 Results for the slowly-changing case

4.6.1 Positive result for mixing in the dense and slowly-changing case

The aim of this section is to prove the following theorem.

Theorem 4.4.3 (Slowly-changing and dense, fast mixing)**.** *Let $d=\Omega(\log n)$, $p=O(\log n/n^2)$, and $q=O(\log n/(dn))$. Let the following assumptions on the starting graph $G_0=(V,E_0)$ be satisfied for large enough constants $c_1,c_2,c_3>0$.*

1. $\deg_0(x)=\Theta(d)$ *for any* $x\in V$*;*
2. $|E_0(S,V\setminus S)|\geq c_2\log n|S|$*, for any* $S\subset V$ *with* $|S|\leq c_1\log n$*;*
3. $\Phi_{G_0}\geq c_3\log d/d$*.*

Then, $t_{\mathrm{mix}}(\mathcal{G}(n,p,q))=O(\log n/\Phi_{G_0}^2)$*.*

We start by proving that, if the three assumptions of Theorem 4.4.3 are satisfied, then, for any $t=O(nd\log n)$, the conductance of G_t is not much worse than the conductance of G_0 (with high probability).

Lemma 4.6.1 (Conductance lower bound)**.** *Let $d=\Omega(\log n)$, $p=O(\log n/n^2)$, and $q=O(\log n/(dn))$. Assume that G_0 satisfies assumptions (1),(2),(3) of Theorem 4.4.3. Then, there exists a constant $c>0$ such that, for any $t=O(nd\log n)$ and any vertex $v\in V$,*

$$\mathbb{P}\left[\deg_t(v)\leq\frac{1}{2}\deg_0(v)\right]=O(n^{-4})$$

and

$$\mathbb{P}\left[\Phi_{G_t}\geq c\cdot\Phi_{G_0}\right]=1-O(n^{-4}).$$

The proof of this lemma proceeds as follows: for any $S\subset V$, when an edge is randomly added or removed from the graph, we show that the probability that $|E_t(S,V\setminus S)|$ increases is usually larger than the probability it decreases. Therefore, we model $|E_t(S,V\setminus S)|$ as a random walk on $\mathbb{N}$ with a bias towards large values of $|E_t(S,V\setminus S)|$, i.e., a *birth-and-death* chain. Using standard argument about birth-and-death chains, we show in Lemma 4.6.1 that it is very unlikely that $|E_t(S,V\setminus S)|$ becomes much smaller than $|E_0(S,V\setminus S)|$. By a similar argument, in Lemma 4.6.3 we also show that the degrees of all nodes in S are approximately the same as their original degrees in G_0. This ensures that the conductance

of a single set S is preserved after $t = O(dn \log n)$ steps. We then use a union bound argument to show that, with high probability, the conductance of the entire graph is preserved. For certain value of d, however, we cannot afford to use an union bound an *all* the possible sets of vertices. To overcome this, we show that we need to apply this union bound only for connected sets S. By carefully bounding the total number of such sets with respect to the maximum degree in G_0, we are able to establish the lemma.

We can now give an outline of the proof of Theorem 4.4.3. The idea is to show that $\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - \mathbf{1}\right\|_{2,\pi_{t+1}}$ is smaller than $\left\|\frac{\mu_t}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$ (unless the latter is already very small). We do this by first relating $\left\|\frac{\mu_t}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$ with $\left\|\frac{\mu_{t+1}}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$. More precisely, we can use Lemma 4.3.1 and Lemma 4.6.1 to show that the latter is smaller than the former by a multiplicative factor that depends on Φ_{G_0}. Then, we bound the difference between $\left\|\frac{\mu_{t+1}}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$ and $\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - \mathbf{1}\right\|_{2,\pi_{t+1}}$. In particular, by exploiting the fact that at each step only $O(\log n)$ random edges can be deleted with high probability, we are able to show that $\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - \mathbf{1}\right\|_{2,\pi_{t+1}}$ is not much larger than $\left\|\frac{\mu_{t+1}}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$. Finally, by putting together all these argument, we show that $\left\|\frac{\mu_t}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$ is monotonically decreasing in t, at least until the walk is mixed. This establishes the theorem.

In the following analysis we need a standard Markov chain called *birth-and-death* chain [58, Chapter 2]. It is a random walk $(Z_t)_{t\in\mathbb{N}}$ on $\mathbb{N}$ whose transition probabilities are position based. Formally, we define

$$\Delta Z_t := Z_{t+1} - Z_t = \begin{cases} +1 & w.p.\ b_{Z_t} \\ 0 & w.p.\ r_{Z_t} = 1 - b_{Z_t} - d_{Z_t}\ , \\ -1 & w.p.\ d_{Z_t} \end{cases} \tag{4.8}$$

where $(b_{Z_t}, r_{Z_t}, d_{Z_t})$ are the probabilities that the position of the walk increases by 1, remains, and decreases by 1. b_{Z_t} and d_{Z_t} are functions depending on the position Z_t of the walk. Furthermore, we also need to define the *hitting time* of a random walk $\tau_b^a := \min\{t \in \mathbb{N} : Z_t = b, Z_0 = a\}$, which is the time the walk requires to hit b starting from $Z_0 = a$.

For a birth-and-death chain moving between 0 and m, the stationary distribution of this Markov chain is given as follows:

$$\pi(k) = \frac{w_k}{\sum_{j=0}^{m} w_j}, \tag{4.9}$$

where $w_k = \prod_{i=1}^{k} \frac{b_{i-1}}{d_i}$ and if all $b_{Z_t} = b'$ and $d_{Z_t} = d'$, $w_k = (b'/d')^k$. The birth-and-death Markov chain is *time-reversible* [58] as defined in Sec. 4.3. Hence $\pi(x)P^t(x,y) = \pi(y)P^t(y,x)$ where P is the transition matrix of the Markov chain and $P^t(x,y)$ is the t-step transition probability from x to y.

To prove Lemma 4.6.1, we first lower bound the number of edges in $E_t(S, V \setminus S)$ and then upper bound the volume of S in G_t for any $t = O(nd \log n)$. The following lemma proves that the number of edges on the boundary of any set S will not decrease to ϵ less than the initial number with high probability.

Lemma 4.6.2 (Lower bounding $|E_t(S, V \setminus S)|$)**.** *Given the assumptions in Lemma 4.6.1, for one set S, let $M_t = |E_t(S, V \setminus S)|$ be the number of edges on the boundary of S in G_t. Let $M_0 = m_0$ be the initial number. Denote $\mathbb{P}[\tau_b^a \leq t]$ the probability that a walk starting from a hits b before t. Then for any constant $\epsilon \in (0, 1)$, we have*

$$\mathbb{P}\left[\tau_{m_T}^{m_0} = O(nd \log n)\right] = \begin{cases} O(n^{-c_1|S|}) & if\ |S| = O(\log n) \\ O\left(\frac{\log d}{d}\right)^{-c_2|S|} & if\ |S| = \omega(\log n) \end{cases},$$

where $m_T = \lfloor (1-\epsilon)m_0 \rfloor$ and $c_1, c_2 \geq 4$ are constants.

Proof. By the settings of the regime and a Chernoff bound, the number of total changes in the graph is upper bounded by $O(\log n)$, and this holds for each of the first n^3 steps with probability $1 - n^{-\Omega(1)}$. Hence in one graph changing step, the quantity $|E_t(S, V \setminus S)|$ changes by at most $O(\log n)$. In the following analysis, we condition on this event being true.

We study the changes of the number of edges in $E_t(S, V \setminus S)$ by modeling it as a birth-and-death Markov chain. As mentioned above, from M_t to M_{t+1}, there are at most $O(\log n)$ possible modifications. If we "unpack" them into single changes, then each change can be regarded as a random change on the number of edges in $E(S, V \setminus S)$. Each change on an edge slot can add an edge if there is no edge, remove an edge if there is an edge, or keep it as it is. Hence we build a birth-and-death chain $(M'_s)_{s \in \mathbb{N}}$ where $M'_0 = M_0 = m_0 = m'_0$ and $m_T = m'_S = \lfloor (1-\epsilon)m_0 \rfloor$. M'_s is still the number of edges in $E(S, V \setminus S)$, but the time stamp represents the number of random changes we apply on the graph. From M_t to M_{t+1}, the graph changing step can contribute $O(\log n)$ such random changes for $(M'_s)_{s \in \mathbb{N}}$. We denote $s(t)$ be the total number of random changes before time t (included).

Hence the transition probability of this birth-and-death chain is then defined as: for all $s \in [s(t), s(t+1)]$,

$$\Delta M'_s := M'_{s+1} - M'_s = \begin{cases} +1 & w.p.\ b_{M'_s} \geq \left(1 - \frac{M_t + O(\log n)}{|S|(n-|S|)}\right) \cdot p \\ 0 & w.p.\ r_{M'_s} = 1 - b_{M'_s} - d_{M'_s} \\ -1 & w.p.\ d_{M'_s} \leq \frac{M_t + O(\log n)}{|S|(n-|S|)} \cdot q \end{cases}$$

When the chain is at the position M'_s and one random change happens, the probability that it causes an edge addition is $\left(1 - \frac{M'_s}{|S|(n-|S|)}\right) \cdot p$ because it first needs to choose an

empty slot in $E(S, V \setminus S)$ with probability $\left(1 - \frac{M'_s}{|S|(n-|S|)}\right)$ and then with probability p it adds an edge. However note that from M_t to M_{t+1}, the random changes are not independent because if one random change picks one edge slot, the other changes caused by the same graph changing step cannot choose the same edge slot again. Hence the actual number of empty slot is in fact lower bounded by $|S|(n-|S|) - M_t - O(\log n)$. Hence the probability $b_{M'_s}$ has the lower bound as shown above. This argument works similarly for $d_{M'_s}$. Due to the assumption that the initial graph has degree $\Omega(\log n)$, these bounds would still remain to be of the same order.

Recall that $p = \Omega\left(\frac{\log n}{n^2}\right)$ and $q = O\left(\frac{\log n}{|E|}\right)$, hence the ratio is

$$\frac{b_{M'_s}}{d_{M'_s}} \geq \frac{(|S|(n-|S|) - M_t + O(\log n)) \cdot p}{(M_t + O(\log n)) \cdot q} = \Omega(|S|).$$

The current regime can give us appropriate constants such that $\frac{b_{M'_s}}{d_{M'_s}} \geq 4$. Even if at the beginning the ratio may not satisfy this condition, there always exists a constant threshold ϵ' such that when M'_s falls below $(1-\epsilon')m'_0$, $\frac{b_{M'_s}}{d_{M'_s}}$ becomes less than 4. Hence w.l.o.g. we can wait until then and assume we start with $\frac{b_{M'_s}}{d_{M'_s}} \geq 4$. Note that m'_0 depends on $|S|$ so we do the following case analysis.

For small sets, $|S| \leq 100 \log n$: By the assumption that the conductance is a constant, we have $M_0 = \Omega(|S| \log n)$. By (4.9) $\pi(m'_S) = O(1/4^{\epsilon m'_0})$ and $\pi(m'_0) \geq 3/4$. By the definition of the reversibility of a Markov chain [58]: for any time t,

$$\pi(m'_0) p_t(m'_0, m'_S) = \pi(m'_S) p_t(m'_S, m'_0).$$

where $p_t(x, y)$ means the t-step transition probability from x to y. Hence

$$p_t(m'_0, m'_S) \leq \pi(m'_S) \frac{p_t(m'_S, m'_0)}{\pi(m'_0)} \leq \frac{4}{3}\pi(m'_S).$$

Then we plug in $\pi(m'_t) = O(1/4^{\epsilon m'_0})$,

$$\begin{aligned}
\mathbb{P}\left[\tau_{m_T}^{m_0} = O(nd \log n)\right] &= \mathbb{P}\left[\tau_{m'_S}^{m'_0} = O(nd \log^2 n)\right] \\
&= \sum_{i=0}^{O(nd \log^2 n)} p_i(m'_0, m'_S) \\
&\leq S \cdot \frac{4}{3}\pi(m'_S) \\
&= O(n^{-c_1|S|}),
\end{aligned}$$

where $c_1 \geq 4$ is a constant.

For larger sized sets, $|S| > 100 \log n$: By our assumption that the conductance is

$\Omega(\log d/d)$. We have $m_0' = \Omega(|S|\log d)$. Then $\frac{b_{M_s'}}{d_{M_s'}}$ becomes $O(d/\log d)$. Hence $\pi(m_S') = O((\log d/d)^{\epsilon m_0'})$ and $\pi(m_0') \geq 1 - O(\log d/d)$. By applying a similar analysis we have

$$\begin{aligned}\mathbb{P}\left[\tau_{m_S'}^{m_0'} = O(nd\log^2 n)\right] &= \sum_{i=0}^{O(n\log^3 n)} p_i(m_0', m_S') \\ &\leq T \cdot \frac{1}{1 - O(\log d/d)} \cdot \pi(m_S') \\ &= O((\log d/d)^{-c_2|S|})\end{aligned}$$

where $c_2 \geq 4$ is a constant. □

Lemma 4.6.3 (Upper bounding the volume, $\mathrm{vol}(S)$)**.** *Given the assumption in Lemma 4.6.1, for one set S, let $N_t = \mathrm{vol}_t(S)$, the sum of the degrees of the vertices of S in G_t. Let $N_0 = n_0$ be the initial volume. Denote by $\mathbb{P}\left[\tau_b^a \leq t\right]$ the probability that a walk starting at a hits b before t. Then for any constant $\delta \geq 0$, we have*

$$\mathbb{P}\left[\tau_{n_T}^{n_0} = O(nd\log n)\right] = O(n^{-c|S|}),$$

where $n_T = \lceil(1+\delta)n_0\rceil$ and $c \geq 4$.

Proof. Denote N_t the volume of the set S at time t then we have another birth-and-death chain $(N_s')_{s\in\mathbb{N}}$ similar to the previous proof where $N_0' = N_0 = n_0$. Let $s(t)$ denote the number of random changes before time t (included). For all $s \in [s(t), s(t+1)]$,

$$\Delta N_s' := N_{s+1}' - N_s' = \begin{cases} 1 & w.p.\ b_{N_s'} \leq \left(1 - \frac{N_t - O(\log n)}{(n-1)|S|}\right) \cdot p \\ 0 & w.p.\ 1 - b_{N_s'} - d_{N_s'} \\ -1 & w.p.\ d_{N_s'} \geq \frac{N_t - O(\log n)}{(n-1)|S|} \cdot q \end{cases}$$

In general this proof is similar to the previous one. The ratio between b and d is

$$\frac{b}{d} \leq O\left(\frac{((n-1)|S| - N_t + O(\log n)) \cdot p}{(N_t - O(\log n)) \cdot q}\right).$$

This ratio is always a constant in our regime. Similar to the previous proof, we may not have a good ratio at the beginning, but there is always a constant threshold δ' such that when $N_s' = (1+\delta')N_0'$ the ratio b/d is less than 1. Then by applying the same birth-and-death chain technique, we prove the lemma. □

Now by combining the above two lemmas we can prove Lemma 4.6.1.

Proof of Lemma 4.6.1. We have showed that for one set S, within $O(n\log^2 n)$ steps, the number of the edges on the boundary will not be ϵ smaller and the volume will not be δ larger than they originally were with high probability.

We apply a union bound to bound the conductance Φ_G of the entire graph. By Lemma 4.9.3, Φ_G is revealed by just looking at the connected sets. Let $\tilde{G}$ be the union of all graphs from $G_1, ..., G_t$. Then all the possible connected sets that ever exist in $\mathcal{G}$ can be found in $\tilde{G}$. The number of all connected sets for one certain graph G_i is upper bounded by those in the union graph. By Lemma 4.9.4, the number of connected sets of size $|S|$ is bounded by $n \cdot \Delta^{|S|}$ where Δ is the maximum degree in $\tilde{G}$. By applying the birth-and-death chain argument for $|S|= 1$, the maximum degree Δ of the union graph should be upper bounded by $O(d)$. Hence below we use $n \cdot O(d^{2|S|-2})$ for the union bound when needed.

Denote $\mathcal{E}_1(S)$ the event that the number of the edges on the boundary of S ever reaches the ϵ less than the beginning. By union bound and Lemma 4.6.2, the probability that there exists such a set is upper bounded by

$$\begin{aligned}
\mathbb{P}\left[\bigcup_{S\subseteq V} \mathcal{E}_1(S)\right] \leq\ & \mathbb{P}\left[\bigcup_{\substack{S\subseteq V:\\ |S|=O(\log n)}} \mathcal{E}(S)\right] + \mathbb{P}\left[\bigcup_{\substack{S\subseteq V:\\ |S|\in[\omega(\log n),O(n/\log n)]}} \mathcal{E}_1(S)\right] \\
& + \mathbb{P}\left[\bigcup_{\substack{S\subseteq V:\\ |S|\in[\omega(n/\log n),O(n)]}} \mathcal{E}_1(S)\right] \\
\leq\ & \sum_{|S|=O(\log n)} n \cdot \Delta^{|S|} \cdot O(n^{-c_1|S|}) \\
& + \sum_{[\omega(\log n),O(n/\log n)]} n \cdot \Delta^{|S|} \cdot O\left(\left(\frac{\log d}{d}\right)^{-c_2|S|}\right) \\
& + \sum_{[\omega(n/\log n),O(n)]} \binom{n}{|S|} O\left(\left(\frac{\log d}{d}\right)^{-c_2|S|}\right) \\
\leq\ & O(n^{-c_3})
\end{aligned}$$

where c_1, c_2 are the constants used in Lemma 4.6.2 and $c_3 \geq 4$ is a constant.

Denote $\mathcal{E}_2(S)$ the event that the volume of S reaches δ larger than the beginning. By union bound and Lemma 4.6.3, the probability that there exists such a set is upper bounded by

$$\mathbb{P}\left[\bigcup_{S\subseteq V} \mathcal{E}_2(S)\right] \leq \sum_{S\subseteq V} \binom{n}{|S|} O(n^{-c|S|}) = O(n^{-c'})$$

where c is the constant used in Lemma 4.6.3 and $c' \geq 4$ is a constant.

By combining everything above, we can lower bound the conductance. If the initial

conductance is lower bounded by ϕ, i.e., $\Phi_G^0(S) = \frac{X_0}{Y_0} \geq \phi$ and

$$\Phi_G^t(S) \geq \frac{(1-\epsilon)X_0}{(1+\delta)Y_0} \geq \frac{(1-\epsilon)\phi}{(1+\delta)},$$

where ϵ, δ are the constants used in Lemma 4.6.2 and Lemma 4.6.3. For simplicity and also because of the arguments we have made along this proof, we choose come constants to be 4 in our final statement of the lemma. □

Proof of Theorem 4.4.3. We establish the theorem by showing unless $\left\|\frac{\mu_t}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$ is already small, $\left\|\frac{\mu_t}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$ will significantly decrease at each step. In particular we relate $\left\|\frac{\mu_t}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$ to $\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - \mathbf{1}\right\|_{2,\pi_{t+1}}$ in two steps:

(1) We lower bound the change between $\left\|\frac{\mu_t}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$ and $\left\|\frac{\mu_{t+1}}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$;

(2) We upper bound the difference between $\left\|\frac{\mu_{t+1}}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$ and $\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - \mathbf{1}\right\|_{2,\pi_{t+1}}$.

Step 1: For the first step, we use a spectral argument. Using Lemma 4.3.1:

$$\left\|\frac{\mu_t}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}^2 - \left\|\frac{\mu_{t+1}}{\pi_t}^2 - \mathbf{1}\right\|_{2,\pi_t}^2 \geq (1 - \lambda_2^2(P_t)) \left\|\frac{\mu_t}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}^2,$$

by rearranging terms we get

$$\left\|\frac{\mu_{t+1}}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}^2 \leq \lambda_2^2(P_t) \left\|\frac{\mu_t}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}^2,$$

where $\lambda_2(P_t)$ is the second largest eigenvalue of P_t, the transition matrix of G_t.
Step 2: Next, we upper bound the difference between $\left\|\frac{\mu_{t+1}}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}$ and $\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - \mathbf{1}\right\|_{2,\pi_{t+1}}$. Due to the randomness of the graph we will compute the expectation of this difference. In the following analysis we condition on the event that at any time t, $|E_t| \in [(1-o(1))nd, (1+o(1))nd]$ where $d = (n-1)\tilde{p}$. This event has probability $1 - o(1)$. Recall that

$$\left\|\frac{\mu_t}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}^2 = \sum_{y \in V} \pi_t(y) \left(\frac{\mu(y)}{\pi_t(y)} - 1\right)^2 = \left(\sum_{y \in V} \frac{\mu_t^2(y)}{\pi_t(y)}\right) - 1.$$

Hence we have

$$\mathbb{E}\left[\left\|\frac{\mu_{t+1}}{\pi_{t+1}} - \mathbf{1}\right\|_{2,\pi_{t+1}}^2 - \left\|\frac{\mu_{t+1}}{\pi_t} - \mathbf{1}\right\|_{2,\pi_t}^2\right]$$
$$= \mathbb{E}\left[\left(\sum_{y \in V} \frac{\mu_{t+1}^2(y)}{\pi_{t+1}(y)}\right) - \left(\sum_{y \in V} \frac{\mu_{t+1}^2(y)}{\pi_t(y)}\right)\right]$$

$$
\begin{aligned}
&= \sum_{y\in V} \mathbb{E}\left[\mu_{t+1}^2(y)\left(\frac{1}{\pi_{t+1}(y)} - \frac{1}{\pi_t(y)}\right)\right] \\
&= \sum_{y\in V} \mathbb{E}\left[\mu_{t+1}^2(y)\left(\frac{2|E_{t+1}|}{\deg_{t+1}(y)} - \frac{2|E_t|}{\deg_t(y)}\right)\right] \\
&\le 2(1+o(1))|E|\sum_{y\in V} \mathbb{E}\left[\mu_{t+1}^2(y)\left(\frac{1}{\deg_{t+1}(y)} - \frac{1}{\deg_t(y)}\right)\right] \\
&\le 2(1+o(1))|E|\sum_{y\in V} \mu_{t+1}^2(y)\mathbb{E}\left[\left(\frac{1}{\deg_{t+1}(y)} - \frac{1}{\deg_t(y)}\right)\right] \qquad (4.10)\\
&\le 2(1+o(1))|E|\sum_{y\in V} \mu_{t+1}^2(y)\frac{\epsilon\deg_t(y)}{\deg_t(y)\cdot(1-\epsilon)\deg_t(y)}(1-(1-q)^{\deg_t(y)}) \qquad (4.11)\\
&\le \frac{2(1+o(1))}{1-o(1)}\sum_{y\in V}\frac{\epsilon}{(1-\epsilon)}\cdot\frac{\mu_{t+1}^2(y)}{\deg_t(y)/((1-o(1))|E|)}(1-(1-q)^{\deg_t(y)}) \\
&\le \frac{2(1+o(1))}{1-o(1)}\cdot\frac{\epsilon}{(1-\epsilon)}(1-(1-q)^{\deg_t(y)})\sum_{y\in V}\frac{\mu_{t+1}^2(y)}{\pi_t(y)} \\
&\le O\left(\frac{\log n}{n}\right)\left(\left\|\frac{\mu_{t+1}}{\pi_t}-\mathbf{1}\right\|_{2,\pi_t}^2+1\right) \qquad (4.12)
\end{aligned}
$$

where $|E|= nd$ and $d = (n-1)\tilde{p}$. The ϵ is the constant used by Lemma 4.6.1. $\deg_{t+1}(y)$ will not decrease to ϵ smaller than $\deg_t(y)$ with probability $1-O(n^{-c})$. From line (4.10) to line (4.11) we upper bound the expectation by only considering the cases where the difference is positive, i.e., $\deg_t(y) \ge \deg_{t+1}(y)$. In line (4.11), by Lemma 4.6.1 we know $\deg_{t+1}(y)$ will not be smaller than $\frac{1}{2}\cdot\deg_t(y)$ with probability $1-O(n^{-4})$. Moreover, the probability $1-(1-q)^{\deg_t(y)}$ is the probability that at least one of the edges connected to y at time t changes at $t+1$. In line (4.12), we hide unimportant constants in the O-notation and we use the inequality $(1-q)^{\deg_t(y)} \ge 1-q\cdot\deg_t(y)$. Since $q = O(\log n/(dn))$ by assumption, we get $O(\log n/n)$ in line (4.12).

By combining the two steps above we have

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\frac{\mu_{t+1}}{\pi_{t+1}}-\mathbf{1}\right\|_{2,\pi_{t+1}}^2 - \left\|\frac{\mu_t}{\pi_t}-\mathbf{1}\right\|_{2,\pi_t}^2\right] \\
&\le O\left(\frac{\log n}{n}\right)\left(\left\|\frac{\mu_{t+1}}{\pi_t}-\mathbf{1}\right\|_{2,\pi_t}^2+1\right) - (1-\lambda_2^2(P_t))\left\|\frac{\mu_t}{\pi_t}-\mathbf{1}\right\|_{2,\pi_t}^2 \\
&\le O\left(\frac{\log n}{n}\right)\left(\lambda_2^2(P_t)\left\|\frac{\mu_t}{\pi_t}-\mathbf{1}\right\|_{2,\pi_t}^2+1\right) - (1-\lambda_2^2(P_t))\left\|\frac{\mu_t}{\pi_t}-\mathbf{1}\right\|_{2,\pi_t}^2 \\
&\le \left(\frac{n+\log n}{n}\cdot\lambda_2^2(P_t)-1\right)\left\|\frac{\mu_t}{\pi_t}-\mathbf{1}\right\|_{2,\pi_t}^2 + O\left(\frac{\log n}{n}\right)
\end{aligned}
$$

Therefore, it holds that

$$\mathbb{E}\left[\left\|\frac{\mu_{t+1}}{\pi_{t+1}}-\mathbf{1}\right\|^2_{2,\pi_{t+1}}\right]\leq\left(\frac{n+\log n}{n}\right)\lambda_2^2(P_t)\cdot\left\|\frac{\mu_t}{\pi_t}-\mathbf{1}\right\|^2_{2,\pi_t}+O\left(\frac{\log n}{n}\right).$$

By Theorem 2.4.2 and the laziness of the walk,

$$\frac{\Phi^2_{G_t}}{2}\leq 1-\lambda_2(P_t)\leq 2\Phi_{G_t}.$$

Since we assume the conductance is lower bounded by $O(\log d/d)$, we have $\lambda_2(P_t)\leq 1-O(\log^2 d/d^2)$ and hence $((n+\log n)/n)\lambda_2^2(P_t)\leq 1$. Therefore in expectation the ℓ_2 distance shrinks by a factor less than 1 but with an additive term. By a similar analysis for the static graph in [58], after $O(\log n/\Phi^2_{G_0})$ rounds, the expected distance to π_t is at most $O(\sqrt{\log n/n})$. By Lemma 4.6.1, we know this holds for $poly(n)$ time. Hence it suffices to apply Markov's inequality and union bound to show that the expected distance is small with probability $1-O(n^{-c})$ on a polynomially long time interval as required in the mixing time notion. □

4.6.2 Negative result for mixing in the sparse and slowly-changing case

Proposition 4.4.4 (Slowly-changing and sparse, slow mixing)**.** *Let $p=O(1/n^2)$ and $q=\omega(1/(n\log n))$. Consider a random walk on $\mathcal{G}(n,p,q)$ with starting graph $G_0\sim\mathcal{G}(n,\tilde{p})$ with $\tilde{p}=p/(p+q)$. Then, $t_{\text{mix}}(\mathcal{G}(n,p,q))=\Omega(n)$.*

Proof. Consider the graph $G_0\sim\mathcal{G}(n,\tilde{p})$. Notice that $\tilde{p}=o(\log n/n)$ is well below the connectivity threshold of Erdős-Rényi random graphs. Therefore, with high probability, there is at least one isolated vertex in G_0; call this vertex u and assume the random walk starts from that vertex. The probability that u remains isolated in the steps $1,2,\ldots,t$ is at least

$$(1-p)^{(n-1)\cdot t}\geq(1-O(1/n^2))^{(n-1)\cdot t}\geq 1-O(t/n).$$

Therefore, with at least constant nonzero probability, there exists a constant $c>0$ such that, for any $t\leq c\cdot n$, $\mu_t^u(u)=1$. Since $\pi_t(u)=0$, this implies that $\|\mu_t^u-\pi_t\|_{TV}=1$. □

Actually the proof reveals a stronger "non-mixing" property; if the random walk starts from a vertex that is isolated in G_0, then this vertex will remain isolated for $\Theta(1/(np))$ rounds in expectation, and in this case the random walk did not move at all!

4.7 Discussion

We now discuss where our results fit in the literature. Classical research studies random walks on static graphs with different topologies (see references in the introductory section). Recently, more and more research has started focusing on dynamic graphs [11, 56, 25, 26, 27, 89, 74, 73, 47, 86].

To be more practical, we intentionally choose a special type of dynamic graph, i.e., the evolving graph model. The main feature of this type is that each graph (except for the first one) depends on the previous graph in the sampled graph sequence, which is closer to the real-world scenarios. Another benefit is that this helps us avoid some unnecessary counter examples in the analysis because an arbitrary dynamic graph sequence can be used by an adversary to create a bad graph to deny many meaningful results.

Previous results, especially in [11, 86], assumed that the graphs in the sequence share the same stationary distribution, which is quite a strong assumption. In our analysis, we do not assume it and classify the model into smaller regimes by controling the parameters (p and q). Since we relax the condition, the result is slightly weaker than previous ones. We show the convergence behaviours in certain regimes (like slowly-changing and dense regime) and some negative results in the other regimes. This is the main difference and contribution we have made in our research.

4.8 Conclusion

In this chapter, we investigated the mixing time of random walks on the edge-Markovian random graph model. Our results cover a wide range of different densities and speeds by which the graph changes. On a high level, these findings provide some evidence of the intuition that both of the two properties "high density" and "slow changes" correlate with fast mixing.

For further work, one interesting setting that is not fully understood is the semi-sparse ($d = \Theta(\log n)$) and fast-changing ($q = \Omega(1) > 0$) case. While we proved that the random walk achieves some coarse mixing in $O(\log n)$, we conjecture that strong mixing is not possible. Another possible direction for future work is, given the bounds on the mixing time at hand, to derive tight bounds on the cover time (i.e., the expected time for a particle to walk through all vertices). Finally, it would be also interesting to study the mixing time in a dynamic random graph model, where not all edge slots are present (similar to the models studied in [56, 47], where the graph at each step is a random subgraph of a possibly sparse network).

4.9 Missing proofs

As mentioned before, the edge-Markovian graph model $\mathcal{G}(n,p,q)$ is similar to an Erdős-Rényi graph $\mathcal{G}(n,\tilde{p})$ where $\tilde{p}=\frac{p}{p+q}$. The mixing time of the graph chain has not been proven formally in previous works. Hence, we provide a proof for the sake of completeness. We remark that since an edge-Markovian model is a time-homogeneous (i.e., static) Markov chain, the classical definition of mixing time (Definition 4.3.4) applies.

Theorem 4.9.1 (Graph chain mixing time)**.** *For an edge-Markovian model $\mathcal{G}(n,p,q)$, the graph distribution converges to the graph distribution of the random graph model $\mathcal{G}(n,\tilde{p})$ where $\tilde{p}=\frac{p}{p+q}$. For any $\epsilon\in(0,1)$, the mixing time of the graph chain $\mathcal{G}(n,p,q)$ is $t_{\text{mix}}(\epsilon)=O\left(\frac{\log(n/\epsilon)}{\log(1/|1-p-q|)}\right)$ for $p+q\neq 1$, and $t_{\text{mix}}(\epsilon)=1$ if $p+q=1$.*

Proof. Every edge slot can be represented by a two-state (close/open) Markov chain with transition matrix

$$P=\begin{pmatrix}1-p & p\\ q & 1-q\end{pmatrix}$$

and stationary distribution $\left(\frac{q}{p+q},\frac{p}{p+q}\right)$. By using standard Markov chain arguments (see, e.g., [58, Chapter 1]), $1-p-q$ is an eigenvalue of the transition matrix. Hence the convergence rate at each step is $1-p-q$, i.e.,

$$\|\mu_{t+1}-\pi\|_{TV}\leq|1-p-q|\|\mu_t-\pi\|_{TV}\,.$$

To see this, we call the two nodes e,w. Then we have

$$\mu_{t+1}(e)-\frac{q}{p+q}=\mu_t(e)(1-p)$$

Therefore, when $p+q\neq 1$, the mixing time $t_{\text{mix}}(\epsilon)$ of this two-state Markov chain is $O\left(\frac{\log(1/\epsilon)}{\log|1-p-q|}\right)$ where $\epsilon<1$. For all the $\binom{n}{2}$ edge slots, the time that all of them mix is $O\left(\frac{\log\binom{n}{2}+\log(1/\epsilon)}{\log|1-p-q|}\right)$. When $p+q=1$, instead, the graph mixes immediately, which confirms the fact that in this regime the graph model is equivalent to a sequence of independent graphs from $\mathcal{G}(n,\tilde{p})$. □

Remark 4.9.2. *Theorem 4.9.1 essentially tells us that, whenever at least one between p and q is large (e.g., $\Omega(1)$), the graph chain quickly converges to $\mathcal{G}(n,\tilde{p})$ with $\tilde{p}=\frac{p}{p+q}$. This suggests that for a fast-changing edge-Markovian model $\mathcal{G}(n,p,q)$ with $q=\Omega(1)$, we can consider w.l.o.g. the starting graph G_0 as sampled from $\mathcal{G}(n,\tilde{p})$.*

The graph chain mixes like a standard Markov chain. This in fact makes us think if we can prove our results from that perspective. Unfortunately the hard technical problem in that direction is that the graph chain is non-reversible, which means many good tools

for reversible chains cannot be applied. Also the graph chain has a huge space compared with n. It should contain all possible graphs with n nodes. Therefore in this thesis, we consider more combinatorial methods to solve the problem. However the mixing property of the graph chain is still an interesting problem, which deserves more research.

4.9.1 Missing proofs in Section 4.6.1

Lemma 4.9.3. *For any graph G, the conductance of G is equal to*

$$\Phi_G = \min_{\substack{S\subset V, \mathrm{vol}(S)\leq \mathrm{vol}(G)/2 \\ S \text{ is connected}}} \Phi_G(S).$$

Proof. By definition the conductance of a graph Φ_G is

$$\Phi_G = \min_{S\subseteq V, |S|\leq n/2} \frac{|E(S, V\setminus S)|}{\mathrm{vol}(S)}$$

Assuming the graph conductance is achieved by a disconnected set D. W.l.o.g. we assume it has two connected components A and B. Then since there are no edges between A and B, we have

$$\Phi_G = \Phi_G(D) = \frac{|E(A, V\setminus A)|+|E(B, V\setminus B)|}{\mathrm{vol}(A)+\mathrm{vol}(B)}$$

Then by a simple inequality,

$$\Phi_G(D) \geq \min\left\{\frac{|E(A, V\setminus A)|}{\mathrm{vol}(A)}, \frac{|E(B, V\setminus B)|}{\mathrm{vol}(B)}\right\}$$

where $|A|, |B|\leq |D|\leq n/2$. By induction, this holds for all disconnected sets with more than two connected components. Hence if there exists a disconnected set D which gives the minimum conductance over the entire graph then there must exist a connected set S in the graph which also achieves this conductance. □

Lemma 4.9.4. *Let G be a graph with n vertices and the maximum degree is Δ. The number of all the connected sets with k vertices is at most $n\cdot\Delta^{2k-2}$.*

Proof. For a fixed $k > 0$, we define an encoding which can represent a path of length $2k$ starting from a certain node. First we label n vertices from 1 to n. Then when enumerating all paths of length $2k$, we first output the label of the starting vertex, then for its neighbours we sort them based on their labels and then use the ranks in the following output.

The output should start with a label which represents the root node which is from 1 to n. Then the following numbers are all less than Δ because each of them only means

the rank of a node among all the neighbours of its predecessors. Hence at most we have $n \cdot \Delta^{2k-2}$ encodings.

There is an injective map between all the connected set of size k with the encoding. For any such set, there is a spanning tree of size k. A DFS traversal of this tree would only use each edge of the spanning tree twice. Hence a not necessarily simple path of length $2(k-1)$ must exist. Our encoding essentially gives all possible paths of length $2(k-1)$. So there must be such an injective map. The number of the strings in our output is $n \cdot \Delta^{2(k-1)}$. Hence the total number of connected sets is upper bounded by that as well. □

Chapter 5

Additional Results for Random Walks on Dynamic Graphs

On the way of the study of the mixing property of the random walks in dynamic graph models, we also developed some other related results, which can help us understand the fluctuation of the random walk distribution better.

The reason why we get these results is that, at the very beginning, we are not sure which the best way is to analyse the behaviour of the distribution of the walk. Since the chain is essentially time inhomogeneous, which means the transition matrix of the random walk is changing over the time, it is natural to think of the measure used in Saloff-Coste et al. [79, 80]: *merging time*. Intuitively speaking, this means that for any two initial distributions μ_0 and μ_0', after running the random walk process for both of them on the same inhomogeneous chain, the distance between the two probability distributions is small. In other words, $\mu_t = \mu_0 \mathbf{P}_1 \mathbf{P}_2 \cdots \mathbf{P}_t$ and $\mu_t' = \mu_0' \mathbf{P}_1 \mathbf{P}_2 \cdots \mathbf{P}_t$ are close to each other. However, it is extremely hard to analyse the case where $\mathbf{P}_i$ and $\mathbf{P}_{i+1}$ has dependencies as defined in the edge-Markovian model. Hence in our study we only analyse the independent case, which corresponds to the edge-Markovian model with $p + q = 1$.

Another natural idea is just to look at the largest probability in the distribution, so we can monitor the amplitude of the entire distribution. If it is upper bounded by a sufficiently small value, then we can tell that even thought the walk distribution does not converge to a stationary distribution, it would not fluctuate very much.

5.1 Merging time: another dynamic mixing

In this section, we present the definition of the merging property. Note that we only consider the independent case from the edge-Markovian model. We then prove that the merging phenomenon does appear in the independent case.

Definition 5.1.1 (Merging times)**.** *Given two starting distribution μ_0 and μ_0', the transition matrices at each time is $\{\mathbf{P}_t\}_{t=0}^n$, the merging time is defined as:*

$$t_{\text{merg}}(\epsilon) := \min\{t \in \mathbb{N} : \|\mu_t - \mu_t'\|_{TV} \leq \epsilon\},$$

where $\mu_t = \mu_0 \prod_{i=1}^t \mathbf{P}_t$ and $\mu_t' = \mu_0' \prod_{i=1}^t \mathbf{P}_t$.

The merging time essentially captures the phenomenon that if we start from two different distributions and going along the same time inhomogeneous Markov chain, they eventually get close to the same distribution. This is not generally true for any time inhomogeneous Markov chain, but we can here prove that it can be true for the chain created by the independent case edge-Markovian model, which is interesting.

Theorem 5.1.2 (Merging time in the independent case)**.** *For the independent edge-Markovian model $\mathcal{G}(n,p,q)$ where $p+q=1$, starting from two different simple random walk distributions μ_0 and μ_0' and $\{\mathbf{P}_t\}_{t=0}^n$ are the transition matrices corresponding to the graphs generated by the edge-Markovian model, the merging time is $O(\log n)$.*

Before moving on to the proof of this theorem, we first need to prove a helping lemma.

Lemma 5.1.3. *Given $\mathcal{G}(n,p,q)$ where $p+q=1$, let μ_t be the distribution of the random walk on $\mathcal{G}$ at time t. Let $d=(n-1)\tilde{p} \geq 2$ be the expected degree where $\tilde{p}=p/(p+q)=p$. For any two initial distributions x_0 and y_0 running on the $\mathcal{G}(n,p,q)$, i.e. $x_{t+1}=x_tP_t, y_{t+1}=y_tP_t$, we have*

$$\mathbb{E}\left[\|x_{t+1}-y_{t+1}\|_2^2\right] \leq \left(\frac{1}{4}+\frac{3}{4}e^{-d}\right)\|x_t-y_t\|_2^2.$$

Proof. We use the 2-norm notation only in this proof to avoid unnecessary coefficients. Below we also use $\Gamma_t(u)$ to represent the set of neighbours of u in G_t.

$$\mathbb{E}\left[\|x_{t+1}-y_{t+1}\|_2^2\right] = \mathbb{E}\left[\sum_{u\in V}(x_{t+1}(u)-y_{t+1}(u))^2\right] = \sum_{u\in V}\mathbb{E}\left[(x_{t+1}(u)-y_{t+1}(u))^2\right]$$
$$= \sum_{u\in V}\left[\mathbb{P}\left[|\Gamma_t(u)|=0\right]\cdot(x_t(u)-y_t(u))^2 + \mathbb{P}\left[|\Gamma_t(u)|\geq 1\right]\cdot\mathbb{E}\left[(x_{t+1}(u)-y_{t+1}(u))^2 \,\middle|\, |\Gamma_t(u)|\geq 1\right]\right].$$

The detailed proof can be found in Sec 5.4. Here we omit it.

$$\leq \sum_{u\in V}\left[e^{-d}\cdot(x_t(u)-y_t(u))^2 + \frac{1}{4}(1-e^{-d})\cdot(x_t(u)-y_t(u))^2\right]$$
$$\leq \left(\frac{1}{4}+\frac{3}{4}e^{-d}\right)\|x^t-y^t\|_2^2,$$

which is the desired result. □

Now we can prove Theorem 5.1.2 by using this lemma.

Proof of Theorem 5.1.2. We apply the previous lemma t times and we will get that after time $t = \Theta(\log n)$ the expected distance between the two distributions is exponentially small. Hence for a time interval we can use Markov's inequality and union bound to prove that in that interval the distance keeps being small with probability $1 - O(n^{-c})$ where c is a constant greater than 1. After the two distributions become the same, their distance keeps being 0 forever.

The decreasing 2-norm does not directly imply the decrease of the total variation. However, if the 2-norm distance decreases to an exponentially small value, the total variation distance cannot be too large. For example, if the 2-norm distance is $1/n^3$, then the total variation distance is at most $1/n$. This gap may look huge but it is acceptable because we can reach $1/n^3$ regardint the 2-norm distance in $O(\log n)$ time. □

5.2 A lower bound for the sparse case

On the way to study the random walk in dynamic graphs, we discover an interesting phenomenon: even if the graph at each step is very sparse (please review Section 4.4 for its definition) and also the walk distribution has become smooth already, it can still get some fluctuation with respect to its ℓ_∞ norm (the maximum entry) of the distribution very quickly. This means that the walk distribution will never smooth for a long time, even if the graph changes slowly. It is interesting because usually for normal random walks on sparse graphs, the fluctuation in the walk distribution is very small and may keep steady for a while. However, here we show that even in the next step, the change in the entire distribution can be huge for sparse dynamic graphs.

Theorem 5.2.1. *For all $t = \Omega(\log n)$ and a constant $0 < c \leq 1/600$, with probability $1 - O((\log n)^{\log\log\log n})$ the ℓ_∞ is at least $\Omega\left(\frac{c\log n}{n\log\log^2 n}\right)$.*

Proof. After $t = \Omega(\log n)$, the graph process converges to $\mathcal{G}(n, \tilde{p})$. The proof of this lemma mainly uses the following lemma. The idea is to show that even at a certain time t the probability distribution μ_t is smooth, with high probability in the next time step, there will be a high probability node created.

Lemma 5.2.2. *Starting at an arbitrary time step $t = \Omega(\log n)$, then with probability $1 - n^{-\epsilon}$ where ϵ is a positive constant smaller than 1, the sampled graph $G_{t-1} \sim \mathcal{G}(n, \tilde{p})$ has the following property: once we fix a vector μ_t whose maximum entry is upper bounded by $\frac{c}{n} \cdot \frac{\log n}{\log\log^2 n}$, then with probability $1 - O(1/(\log n)^{\log\log\log n})$ over the randomness of G_t given G_{t-1}, we have $\Omega(n^{0.8})$ nodes with degrees $\Omega\left(\frac{0.1\log n}{\log\log n}\right)$ in G_t and they hold probability mass at least $\frac{c}{n} \cdot \frac{\log n}{\log\log^2 n}$ in μ_{t+1} after one step random walk in G_t where c is a constant such that $0 < c \leq 1/600$.*

The proof of the above lemma can be found after this proof. The lemma says that if we do not have large probabilities in the vector μ_t, i.e., the distribution is smooth, then with high probability in μ_{t+1} we have at least $\Omega(n^{0.8})$ nodes with large probabilities at least $\Omega\left(\frac{\log n}{n \log\log^2 n}\right)$.

□

Furthermore, please note that we have shown that there are not only constant many such nodes but $\Omega(n^{0.8})$ such nodes. This means the distribution can be far away from being uniform or smooth with high probability even if in the previous step it is already very smooth. We can derive a quick corollary from the previous theorem.

Corollary 5.2.3. *For all $t = \Omega(\log n)$, the total variation distance between μ_t and a uniform distribution is lower bounded by $\Omega\left(\frac{\log n}{n^{1-\epsilon}\log\log^2 n}\right)$ where $\epsilon \in (0, 0.8)$ with probability $1 - O((\log n)^{\log\log\log n})$.*

Below we complete the above proof by filling in the details of Lemma 5.2.2.

Proof of Lemma 5.2.2. In this proof we set precise thresholds, a "high load node" holds probability larger than or equal to $c \cdot \frac{\log n}{n \log\log^2 n}$ and a "low load node" has probability strictly less than $c \cdot \frac{\log n}{n \log\log^2 n}$, where c is a constant in $[0, 1/600]$. Similarly, a "high degree node" has more than or equal to $\frac{\log n}{\log\log n}$ neighbours. A "low degree node" has strictly less than $\frac{\log n}{\log\log n}$ neighbours.

Notice that the nodes with very low load cannot contribute much no matter what degrees they have. From this point on we only consider the set of nodes with sufficiently large loads in μ_t. We only focus on the ones that get degrees below $\log\log n$ in G_t conditioning on G_{t-1}. We define a set S_t of nodes each of which has probability at least $\frac{0.1}{n}$ in μ_t.

$$S_t = \left\{ x \in V : \mu_t(x) \geq \frac{0.1}{n} \right\},$$

and $\mu_t(S_t) = \sum_{x \in S_t} \mu_t(x) \geq 0.9$ because $\mu_t(V \setminus S_t) \leq n \cdot \frac{0.1}{n} = 0.1$. Since there are no high load nodes at time t, the size of S_t can be lower bounded:

$$|S_t| \geq n \cdot \frac{0.9 \log\log^2 n}{c \log n}.$$

Denote $\deg_t(x)$ the degree of the node x in G_t and $\deg_t^S(x)$ the number of neighbours of x in the set $S \subseteq V$. The proof sketch is:

1. We extract a subset $S'_t \subseteq S_t$ in which all nodes have degrees less or equal than $\log\log n$ in G_{t-1} because the nodes with degrees larger than $\log\log n$ are very likely to keep most of their neighbours and contribute less in the next step. Formally:

$$S'_t = \left\{ x \in S_t : \deg_{t-1}(x) \leq \log\log n \right\}.$$

2. We only expose the edges inside S'_t in G_t conditioning on the fact that they have less than $\log\log n$ neighbours in G_{t-1}. After the exposure, we remove the nodes with degrees more than $2\log\log n$ within S'_t and get a subset S''_t of S'_t. Formally:

$$S''_t = \left\{ x \in S'_t : \deg_t^{S'_t}(x) \leq 2\log\log n \right\}.$$

3. We prove that in G_t there is at least one node with $\frac{0.1\log n}{\log\log n}$ neighbours in S''_t.

Step 1 Since $t = \Omega(\log n)$, G_{t-1} can be seen as a sample from $\mathcal{G}(n, \tilde{p})$ where $\tilde{p} = \frac{p}{p+q}$. We apply Lemma 5.5.1 on G_{t-1} . By Lemma 5.5.1 and picking the constant in that lemma $\delta = 0.9$, then there are at least $(1-1/\log n)|S_t|$ nodes in S_t with degrees less than $\log\log n$ in G_{t-1}. Taking them out would give us S'_t and $|S'_t| \geq (1-1/\log n)|S_t|$ with probability $1 - O\left(\frac{1}{n}\right)$.

Step 2 Here we only expose the randomness of the edges $\{u, v\}$ for all $u, v \in S'_t$ for G_t. If the node already gets more than $2\log\log n$ edges within S'_t in G_t, we abandon them and get S''_t. We apply a similar argument as Lemma 5.5.1. By union bound, the probability that a node in S'_t has degree more than $2\log\log n$ in S'_t with G_t conditioning on G_{t-1} is:

$$\begin{aligned}
&\forall x \in S'_t,\ \mathbb{P}\left[\deg_t^{S'_t}(x) \geq 2\log\log n \mid G^{t-1}\right] \\
&\leq \sum_{k=2\log\log n}^{n} \sum_{i=0}^{\deg_{t-1}(x)} \binom{|S'_t| - \deg_{t-1}(x)}{k-i} p^{k-i} \binom{\deg_{t-1}(x)}{i} (1-q)^i
\end{aligned}$$

where i is the number of edges kept from the previous graph

$$\begin{aligned}
&\leq \sum_{k=2\log\log n}^{n} \sum_{i=0}^{\deg_{t-1}(x)} \binom{|S'_t| - \deg_{t-1}(x)}{k-i} \left(\frac{1}{n}\right)^{k-i} \binom{\deg_{t-1}(x)}{i} \left(\frac{1}{2}\right)^i \\
&\leq \sum_{k=2\log\log n}^{n} \deg_{t-1}(x) \cdot \max_{0\leq i\leq \deg_{t-1}(x)} \left(\binom{|S'_t| - \deg_{t-1}(x)}{k-i} \left(\frac{1}{n}\right)^{k-i} \binom{\deg_{t-1}(x)}{i} \left(\frac{1}{2}\right)^i \right) \\
&\leq \sum_{k=2\log\log n}^{n} \log\log n \cdot \max_{1\leq i\leq \log\log n} \left(\left(\frac{e|S'_t|}{k-i}\right)^{k-i} \left(\frac{1}{n}\right)^{k-i} \left(\frac{e\log\log n}{i}\right)^i \left(\frac{1}{2}\right)^i \right) \\
&\lesssim \sum_{k=2\log\log n}^{n} \log\log n \cdot \max_{1\leq i\leq \log\log n} \left(\left(\frac{1}{k-i}\right)^{k-i} \left(\frac{\log\log n}{i}\right)^i \right) \\
&\leq \sum_{k=2\log\log n}^{n} \log\log n \cdot \max_{1\leq i\leq \log\log n} \left(\left(\frac{1}{k-i}\right)^{k-i} \sqrt{\log n} \right) \\
&\lesssim \sum_{k=2\log\log n}^{\log n} \left(\frac{1}{2\log\log n - \log\log n}\right)^{2\log\log n - \log\log n} + \sum_{k=\log n}^{n} \left(\frac{1}{\log n - \log\log n}\right)^{\log n - \log\log n}
\end{aligned}$$

$$\leq \log n \cdot \frac{1}{(\log n)^{\log\log\log n}} + n \cdot \frac{1}{n^{\log\log n}}$$
$$\lesssim O\left(\frac{1}{(\log n)^{\log\log\log n}}\right).$$

Note that since $k \geq 2\log\log n$, $\left(\frac{1}{k-i}\right)^{k-i}$ is maximized when i is maximized. Hence the maximum probability here is $O\left(\frac{1}{(\log n)^{\log\log\log n}}\right)$. The derivative of $\left(\frac{\log\log n}{i}\right)^i$ is

$$\frac{\log\log^i n \cdot (\log\log\log n - \log i - 1)}{i^i}.$$

So the zero point is when $i = \frac{\log\log n}{2}$ hence the maximum of it is $\sqrt{\log n}$. Again we apply Chebyshev's inequality as in Lemma 5.5.1:

$$\mathrm{Var}\left[\left|\{x \in S'_t : \deg_t^{S'_t} \geq 2\log\log n\}\right|\right] = O(|S'_t|),$$

and

$$\mathbb{E}\left[\left|\{x \in S'_t : \deg_t^{S'_t} \geq 2\log\log n\}\right|\right] = O\left(\frac{|S'_t|}{(\log n)^{\log\log\log n}}\right),$$

$$\mathbb{P}\left[\left|\{x \in S'_t : \deg_t^{S'_t} \geq 2\log\log n\}\right| \geq \frac{|S'_t|}{\log n} \,\middle|\, G^{t-1}\right] = O\left(\frac{\log^3 n}{n}\right).$$

Hence we take $S''_t \subseteq S'_t$ such that $S''_t = \left\{x \in S'_t : \deg_t^{S'_t}(x) \leq 2\log\log n\right\}$ and $|S''_t| \geq |S'_t| \cdot (1 - 1/\log n)$.

Step 3 For technical reasons we need to restrict the size of S_t to $n/2$. If it is less than $n/2$ then we do nothing. If it is greater, then we simply move the nodes with smaller probabilities to $V \setminus S_t$. Now we design an algorithm to expose the randomness little by little and find a set H such that $\deg_t^{S''_t}(x) \geq \frac{0.1\log n}{\log\log n}$ for all $x \in H$.

To prove this algorithm succeeds with high probability we need to prove the following things:

1. (Line 7-13) With probability $1 - O(\exp(-n^{0.8}))$ the algorithm finds at least $\Omega(n^{0.8})$ nodes in $V \setminus S_t$ such that their degrees in S_t are greater than $\frac{0.1\log n}{\log\log n}$.
2. (Line 16-22) With probability $1 - O(\exp(-n^{0.8}))$ there are at least $|H|/2$ nodes in H such that all their neighbours have degree less than $3\log\log n + 1$.

Proof of point 1: Let $k = \frac{0.1\log n}{\log\log n}$, the probability that u_i has at least k neighbours in L_i is:

$$\mathbb{P}\left[\deg_t^{L_i}(u_i) \geq k\right] \geq \binom{|L_i|}{k}\left(\frac{1}{n}\right)^k\left(1 - \frac{1}{n}\right)^{|L_i|-k}.$$

Algorithm 1 FindHighProbabilityNodes

Input: $S_t^{''}$, $V \setminus S_t$, G_{t-1}.

Output: A set H of nodes with degree at least $\frac{0.1 \log n}{\log \log n}$ in G_t and all their neighbours have degrees less than $3 \log \log n + 1$.

begin
  $H \longleftarrow \emptyset$ $L_1 \longleftarrow S_t^{''}$ **for** $i = 1$ **to** $|V \setminus S_t|$ **do**
    Consider the node $u_i \in V \setminus S_t$ $\deg_t^{L_i}(u_i) \longleftarrow 0$ **for** $j = 1$ **to** $|L_i|$ **do**
      Consider the node $w_j \in L_i$ Reveal the edge $\{w_j, u_i\}$ for G^t **if** $\{w_j, u_i\} \in G^t$ **then**
        $\deg_t^{L_i}(u_i) \longleftarrow \deg_t^{L_i}(u_i) + 1$
      if $\deg_t^{L_i}(u_i) \geq \frac{0.1 \log n}{\log \log n}$ **then**
        break
    Let $\{\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_{\deg_t^{L_i}(u_i)}\}$ be the neighbours of u_i in L_i $T \longleftarrow true$ **for** $k = 1$ **to** $|\deg_t^{L_i}(u_i)|$ **do**
      if $\deg_{V \setminus S_t}^t(\tilde{w}_k) > \log \log n$ **then**
        $T \longleftarrow false$ break
    if T *is true* **then**
      $H \longleftarrow H \cup \{u_i\}$
    $L_{i+1} \longleftarrow L_i \setminus \{\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_{\deg_t^{L_i}(u_i)}\}$ **if** $|L_{i+1}| < |S_t^{''}|/2$ **then**
      break
  return H

In the for loop, each round can remove at most k nodes from L_i. Hence there are at least $|S_t''|/k$ nodes in $V \setminus S_t$ to be checked. To prove 1, we use the second moment method. We define a random variable $|H| = |\{u_i \in V \setminus S_t : \deg_t^{L_i}(u_i) \geq k\}|$. The expectation of X is:

$$\mathbb{E}\left[|H|\right] \geq \frac{|S_t''|}{k} \min_i \left(\binom{|L_i|}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{|L_i|-k} \right)$$

since $|S_t''| = \Omega(n \log\log n/k)$ and $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ and $|L_i| = |S_t''|/2$ minimizes the formula

$$\begin{aligned}
&\gtrsim n \cdot \left(\frac{\log\log n}{k^2}\right)^k \exp\left(-\frac{|L_i|-k}{n}\right) \\
&\gtrsim \exp\left(\log n + k\log\log\log n - 2k\log k - \frac{\log\log n}{k} + \frac{k}{n}\right) \\
&\gtrsim \exp\left(\log n + k(\log\log\log n - 2\log k) + O(\log\log n)\right) \\
&\gtrsim \exp\left(\log n - \frac{0.1\log n}{\log\log n}(2\log\log n + 2\log 0.1 - 2\log\log\log n) + O(\log\log n)\right) \\
&= \Omega(n^{0.8}).
\end{aligned}$$

We apply Chernoff bound here because $\mathbb{1}_{\deg_t^{L_i}(u_i) \geq k}$ for all u_i's are independent. Hence

$$\mathbb{P}\left[|H| \leq \frac{1}{2}\mathbb{E}\left[|H|\right]\right] \leq \exp\left(-\frac{1}{8}\mathbb{E}\left[|H|\right]\right) \leq \exp\left(-\frac{1}{8}\Omega(n^{0.8})\right).$$

Therefore with probability $1 - O(\exp(-n^{0.8}))$, there exists $\Omega(n^{0.8})$ vertex in $V \setminus S_t$ with degree $\frac{0.1\log n}{\log\log n}$.

Proof of point 2: For all the neighbours of u_i, $\{\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_{\deg_t^{L_i}(u_i)}\}$, for any $\tilde{w}_j$, $\deg_t^{V\setminus S_t}(\tilde{w}_j) \leq \log\log n$. In other words, for all $\tilde{w}_j$, $\deg^t(\tilde{w}_j) \leq 3\log\log n$. By a similar argument as above, we have

$$\mathbb{P}\left[\bigcup_{j=1}^{\deg_t^{L_i}(u_i)} \deg_t^{V\setminus S_t}(\tilde{w}_j) \geq \log\log n + 1\right] = O\left(\frac{1}{(\log n)^{\log\log\log n}}\right).$$

Hence

$$\mathbb{P}\left[\bigcap_{j=1}^{\deg_t^{L_i}(u_i)} \deg_t^{V\setminus S_t}(\tilde{w}_j) \leq \log\log n + 1\right] \geq 1 - O\left(\frac{1}{(\log n)^{\log\log\log n}}\right).$$

In conclusion, there are $An^{0.8}$ high degree nodes in H where A is a constant. For each of them the probability that all its neighbours have degree less than $3\log\log n + 1$ is at least $1 - O(1/(\log n)^{\log\log\log n})$. We define a random variable $Y := |\{u_i \in H : \forall \tilde{w}_j \in \Gamma(u_i), \deg_t(w_j) \leq 3\log\log n + 1\}|$. $\mathbb{E}\left[Y\right] \geq An^{0.8}(1 - O(1/(\log n)^{\log\log\log n}))$. By applying

Chernoff again we can have

$$\mathbb{P}\left[Y \leq \frac{\mathbb{E}[Y]}{2}\right] \leq \exp\left(-\frac{1}{8}\mathbb{E}[Y]\right) = \exp\left(-\frac{1}{8}\Omega(n^{0.8})\right).$$

For all the nodes in H, they have at least $\frac{0.1\log n}{\log\log n}$ neighbours each of which has load at least $0.1/n$. Hence each would contribute at least $\frac{1}{60n\log\log n}$ considering the laziness. Then the lower bound of the loads of the nodes in H is $\frac{\log n}{600n\log\log^2 n}$.

In sum, the total probability of this big event is restricted by step 2, hence the overall probability bound should be $1 - O(1/(\log n)^{\log\log\log n})$. □

5.3 Discussion and Experiments

In sum, the merging or the lower bound results both show some interesting properties of the random walk on the edge-Markovian model. The merging result shows that if the graphs are independent from each other, then no matter where you start the random walk, the walk distribution will eventually reach the same distribution after $O(\log n)$ time. The lower bound result shows that the probability distribution will fluctuate quickly even if the underlying graph changes slowly.

We support these observations by some simulations and the results are presented below. We can see that the simulations verified our theoretical results nicely.

5.3.1 Simulation for the merging result

In our simulation (Figure 5.1), the graph has 1000 nodes. The random walk is simple, i.e., without any laziness. Here $p = 1/n = 0.001$ and $q = 0.5$. The graph essentially is very sparse and also changes fast. The starting distributions μ_0 and μ_0' are two delta distributions with different positions holding 1 (hence the initial distance is 2). We can see that after 10 steps (which is roughly $\log n$) the distance become very small, which verifies our result.

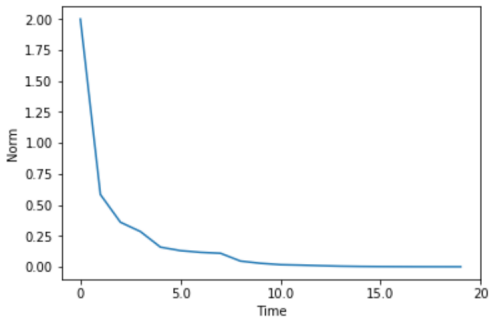


Figure 5.1: Simulation for merging

5.3.2 Simulation for the lower bound result

In this simulation (Figure 5.3), the graph has 1000 nodes again. This time we observe the change of the fluctuation of the ℓ_∞ norm of the distribution. We can see that it decreases first and then fluctuate around a certain value. It never falls below some threshold from the experimental data. The figure may not show the fluctuation clearly but the data fluctuate in $[0.0025, 0.0041]$. The data below was computed by averaging *10 independent runs*.

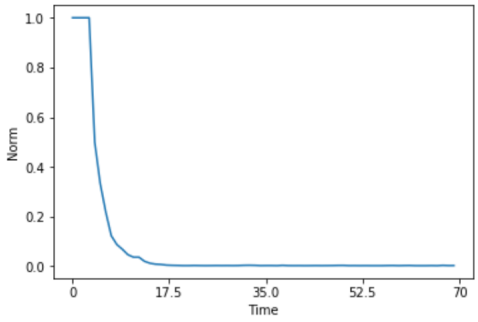


Figure 5.2: Simulation for the ℓ_∞ norm of μ_t

As also mentioned in Corollary 5.2.3, another way to amplify the fluctuation for us to observe is to check the total variation distance between the distribution μ_t and the uniform distribution. Here we use the same data but this time we do not only check the ℓ_∞ norm of it to verify this corollary. We compute the total variation distance between it and the uniform distribution.

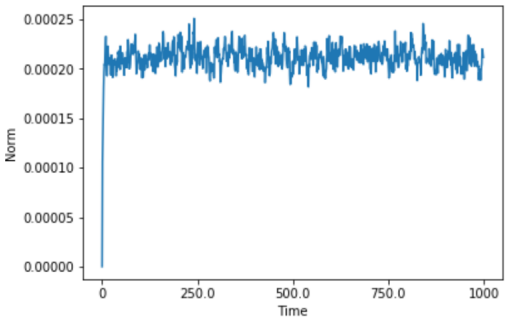


Figure 5.3: The total variation distance between μ_t and the uniform distribution

We can see that the fluctuation never stops but it always stays within a range. The value becomes smaller than the previously mentioned interval because in the total variation distance each entry of μ_t needs to subtract $1/n$.

5.4 Missing proofs

5.4.1 Missing steps in Lemma 5.1.3

Here we show the missing steps in the previous section. Due to the technical reasons, we use a top-down style to present it, i.e., we first demonstrate the main steps and leave the complicated calculation to later subsections.

$$
\begin{aligned}
&\mathbb{E}\left[\|x_{t+1}-y_{t+1}\|_2^2\right] = \mathbb{E}\left[\sum_{u\in V}(x_{t+1}(u)-y_{t+1}(u))^2\right] = \sum_{u\in V}\mathbb{E}\left[(x_{t+1}(u)-y_{t+1}(u))^2\right] \\
&= \sum_{u\in V}\left[\mathbb{P}\left[|\Gamma_t(u)|=0\right]\cdot(x_t(u)-y_t(u))^2 + \mathbb{P}\left[|\Gamma_t(u)|\neq 0\right]\cdot\mathbb{E}\left[(x_{t+1}(u)-y_{t+1}(u))^2 \mid |\Gamma_t(u)|\neq 0\right]\right] \\
&= \sum_{u\in V}\Bigg[\mathbb{P}\left[|\Gamma_t(u)|=0\right]\cdot(x_t(u)-y_t(u))^2 \\
&\qquad +\mathbb{P}\left[|\Gamma_t(u)|\neq 0\right]\cdot\mathbb{E}\left[\left(\frac{1}{2}(x_t(u)-y_t(u)) + \frac{1}{2}\sum_{v\in\Gamma_t(u)}\frac{x_t(v)-y_t(v)}{\deg_t(v)}\right)^2 \,\middle|\, |\Gamma_t(u)|\neq 0\right]\Bigg] \\
&= \sum_{u\in V}\Bigg[\left(1-\frac{d}{n}\right)^n\cdot(x_t(u)-y_t(u))^2 \\
&\qquad +\left(1-\left(1-\frac{d}{n}\right)^n\right)\cdot\mathbb{E}\left[\left(\frac{1}{2}(x_t(u)-y_t(u)) + \frac{1}{2}\sum_{v\in\Gamma_t(u)}\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2 \,\middle|\, |\Gamma_t(u)|\neq 0\right]\Bigg]
\end{aligned}
\tag{5.1}
$$

We expand this expectation later.

$$
\leq \left(\frac{1}{4}+\frac{3}{4}e^{-d}\right)\|x_t-y_t\|_2^2,
$$

where Z_v is the random variable for the number of neighbours of a vertex v except u after we have known that v is a neighbour of u. Note that for $v_1, v_2 \in \Gamma_t(u)$, Z_{v_1}, Z_{v_2} are negatively correlated. $|\Gamma_t(u)|$ is a binomial random variable from $\mathsf{Bin}(n-1,\tilde{p})$ and $\frac{1}{1+Z_v}$ is also a random variable.

5.4.2 Expansion of Line (5.1)

Below is the expansion step, we omitted the condition on $\deg_t(u) \neq 0$ to simplify the expression.

$$
\mathbb{E}\left[\left(\frac{1}{2}(x_t(u)-y_t(u)) + \frac{1}{2}\sum_{v\in\Gamma_t(u)}\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2\right]
$$

$$= \mathbb{E}\left[\frac{1}{4}(x_t(u)-y_t(u))^2 + \frac{1}{2}(x_t(u)-y_t(u))\sum_{v\in\Gamma_t(u)}\frac{x_t(v)-y_t(v)}{1+Z_v} + \frac{1}{4}\left(\sum_{v\in\Gamma_t(u)}\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2\right]$$

$$= \frac{1}{4}(x_t(u)-y_t(u))^2 + \frac{1}{2}(x_t(u)-y_t(u))\mathbb{E}\left[\sum_{v\in\Gamma_t(u)}\frac{x_t(v)-y_t(v)}{1+Z_v}\right] \tag{5.2}$$

$$+ \frac{1}{4}\mathbb{E}\left[\left(\sum_{v\in\Gamma_t(u)}\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2\right] \tag{5.3}$$

$$\leq \frac{1}{4}(x_t(u)-y_t(u))^2 + \frac{1}{2}(x_t(u)-y_t(u))\left(\frac{y_t(u)-x_t(u)}{n-1}\right)\frac{1-4^{-d}}{1-e^{-d}}$$

$$+ \frac{1}{4}\left(\frac{d}{1-e^{-d}}\frac{\|x_t-y_t\|_2^2-(x_u-y_u)^2}{n-1}\frac{2-2e^{-d}d-2e^{-d}}{d^2}\right.$$

$$\left. + \frac{d(d-\tilde{p})}{1-e^{-d}}\left(\frac{y_t(u)-x_t(u)}{n-1}\right)^2\left(\frac{1-4^{-d}}{d}\right)^2\right)$$

5.4.3 Expansion of the numbered lines

Now we analyse (5.2) and (5.3):

$$(5.2): \mathbb{E}\left[\sum_{v\in\Gamma_t(u)}\frac{x_t(v)-y_t(v)}{1+Z_v}\right]$$

$$= \sum_{i=1}^{n-1}\mathbb{P}\left[|\Gamma_t(u)|= i \,\middle|\, |\Gamma_t(u)|\neq 0\right]\cdot\mathbb{E}\left[\sum_{v\in\Gamma_t(u)}\frac{x_t(v)-y_t(v)}{1+Z_v}\,\middle|\,|\Gamma_t(u)|= i\right]$$

We expose the randomness of the graph for all Z_vs, and all choices of the $x_t(v), y_t(v)$

$$= \sum_{i=1}^{n-1}\mathbb{P}\left[|\Gamma_t(u)|= i \,\middle|\, |\Gamma_t(u)|\neq 0\right]$$

$$\cdot \sum_{Z_{v_1},...,Z_{v_i}}\mathbb{P}\left[Z_{v_1}=z_1,...,Z_{v_i}=z_i\right]\sum_{v_1,...,v_i}\mathbb{P}\left[v_1=w_1,...,v_i=w_i\right]\sum_{j=1}^{i}\frac{x_t(w_j)-y_t(w_j)}{1+Z_{v_j}}$$

$$= \sum_{i=1}^{n-1}\mathbb{P}\left[|\Gamma_t(u)|= i \,\middle|\, |\Gamma_t(u)|\neq 0\right]$$

$$\cdot \sum_{Z_{v_1},...,Z_{v_i}}\sum_{v_1,...,v_i}\mathbb{P}\left[Z_{v_1}=z_1,...,Z_{v_i}=z_i\right]\mathbb{P}\left[v_1=w_1,...,v_i=w_i\right]\sum_{j=1}^{i}\frac{x_t(w_j)-y_t(w_j)}{1+Z_{v_j}}$$

For the permutations σ of $\mathbf{Z}$ by symmetry we know we always get the sum of all the nodes except u

$$
\begin{aligned}
&= \sum_{i=1}^{n-1} \mathbb{P}\left[|\Gamma_t(u)| = i \,\middle|\, |\Gamma_t(u)| \neq 0 \right] \\
&\quad \cdot \sum_{Z_{v_1},...,Z_{v_i}} \sum_{v_1,...,v_i} \mathbb{P}\left[(Z_{v_1},...,Z_{v_i}) \in \sigma(z_1,...,z_i) \right] \mathbb{P}\left[v_1 = w_1, ..., v_i = w_i \right] \\
&\quad \cdot \sum_{j=1}^{i} \frac{(i-1)!}{1+z_j} \sum_{k=1}^{i} (x_t(w_k) - y_t(w_k))
\end{aligned}
$$

Assume we pick $\mathbf{v}$ following an order of the vertices, then they have the same probability $1/\binom{n-1}{i}$ to be chosen. For each $x_t(v)$, after we sum all combination, it is chosen for $\binom{n-1}{i}/(n-1)$ times.

$$
\begin{aligned}
&= \sum_{i=1}^{n-1} \mathbb{P}\left[|\Gamma_t(u)| = i \,\middle|\, |\Gamma_t(u)| \neq 0 \right] \\
&\quad \cdot \sum_{Z_{v_1},...,Z_{v_i}} \mathbb{P}\left[(Z_{v_1},...,Z_{v_i}) \in \sigma(z_1,...,z_i) \right] \frac{1}{\binom{n-1}{i}} \sum_{j=1}^{i} \frac{(i-1)!}{1+z_j} \frac{\binom{n-1}{i}}{n-1} (y_t(u) - x_t(u)) \\
&= \left(\frac{y_t(u) - x_t(u)}{n-1} \right) \sum_{i=1}^{n-1} \mathbb{P}\left[|\Gamma_t(u)| = i \,\middle|\, |\Gamma_t(u)| \neq 0 \right] \\
&\quad \cdot \sum_{Z_{v_1},...,Z_{v_i}} \mathbb{P}\left[(Z_{v_1},...,Z_{v_i}) \in \sigma(z_1,...,z_i) \right] \sum_{j=1}^{i} \frac{(i-1)!}{1+z_j}
\end{aligned}
$$

Due to the negative correlation proved in Proposition 5.4.1, $\mathbb{P}\left[Z_{v_1}, Z_{v_2} \right] \leq \mathbb{P}\left[Z_{v_1} \right] \mathbb{P}\left[Z_{v_2} \right]$

$$
\begin{aligned}
&\leq \left(\frac{y_t(u) - x_t(u)}{n-1} \right) \sum_{i=1}^{n-1} \mathbb{P}\left[|\Gamma_t(u)| = i \,\middle|\, |\Gamma_t(u)| \neq 0 \right] \cdot i \cdot \mathbb{E}\left[\frac{1}{1+Z_v} \right] \qquad (5.4) \\
&= \left(\frac{y_t(u) - x_t(u)}{n-1} \right) \mathbb{E}\left[|\Gamma_t(u)| \,\middle|\, |\Gamma_t(u)| \neq 0 \right] \cdot \mathbb{E}\left[\frac{1}{1+Z_v} \right] \\
&\leq \left(\frac{y_t(u) - x_t(u)}{n-1} \right) \frac{d}{1-e^{-d}} \frac{1-4^{-d}}{d} = \left(\frac{y_t(u) - x_t(u)}{n-1} \right) \frac{1-4^{-d}}{1-e^{-d}}
\end{aligned}
$$

In the above and below calculations, (5.4) is used many times. It involves the expectation of the reciprocal of the random variable Z_v. The details on how to compute this quantity is delayed to Section 5.4.4.2.

Second, we analyse the quadratic term (5.3).

$$
(5.3) : \mathbb{E}\left[\left(\sum_{v \in \Gamma_t(u)} \frac{x_t(v) - y_t(v)}{1+Z_v} \right)^2 \right]
$$

$$
\begin{aligned}
&\leq \mathbb{E}\left[|\Gamma_t(u)| \,\middle|\, |\Gamma_t(u)|\neq 0\right] \mathbb{E}\left[\left(\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2\right] \\
&\quad + \mathbb{E}\left[|\Gamma_t(u)|^2 - |\Gamma_t(u)| \,\middle|\, |\Gamma_t(u)|\neq 0\right] \mathbb{E}\left[\frac{x_t(v)-y_t(v)}{1+Z_v}\right]^2 \\
&\leq \frac{(n-1)\tilde{p}}{1-(1-\tilde{p})^{n-1}} \mathbb{E}\left[\left(\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2\right] + \frac{(n-1)^2\tilde{p}^2-(n-1)\tilde{p}^2}{1-(1-\tilde{p})^{(n-1)}} \mathbb{E}\left[\frac{x_t(v)-y_t(v)}{1+Z_v}\right]^2 \\
&\leq \frac{d}{1-e^{-d}} \mathbb{E}\left[\left(\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2\right] + \frac{d(d-\tilde{p})}{1-e^{-d}} \mathbb{E}\left[\frac{x_t(v)-y_t(v)}{1+Z_v}\right]^2 \\
&\leq \frac{d}{1-e^{-d}} \mathbb{E}\left[(x_t(v)-y_t(v))^2\right] \mathbb{E}\left[\left(\frac{1}{1+Z_v}\right)^2\right] + \frac{d(d-\tilde{p})}{1-e^{-d}} \mathbb{E}\left[x_t(v)-y_t(v)\right]^2 \mathbb{E}\left[\frac{1}{1+Z_v}\right]^2 \qquad (5.5) \\
&\leq \frac{d}{1-e^{-d}} \frac{\|x_t-y_t\|_2^2-(x_u-y_u)^2}{n-1} \mathbb{E}\left[\left(\frac{1}{1+Z_v}\right)^2\right] + \frac{d(d-\tilde{p})}{1-e^{-d}} \left(\frac{y_t(u)-x_t(u)}{n-1}\right)^2 \mathbb{E}\left[\frac{1}{1+Z_v}\right]^2 \\
&\leq \frac{d}{1-e^{-d}} \frac{\|x_t-y_t\|_2^2-(x_u-y_u)^2}{n-1} \frac{2-2e^{-d}d-2e^{-d}}{d^2} + \frac{d(d-\tilde{p})}{1-e^{-d}} \left(\frac{y_t(u)-x_t(u)}{n-1}\right)^2 \left(\frac{1-4^{-d}}{d}\right)^2.
\end{aligned}
$$

Here (5.5) is similar to (5.4). We leave the calculation of it in Section 5.4.4.2.

5.4.4 Missing proofs in the previous section

5.4.4.1 Negative correlation

We need to use a concept in probability in the above proof: negative correlation. X, Y are called negatively correlated if $\mathbb{P}[X,Y] \leq \mathbb{P}[X] \cdot \mathbb{P}[Y]$.

Proposition 5.4.1. *Let Z_v be the random variable which represents the number of neighbours of a vertex v except u after we have known that v is a neighbour of u. We prove the negative correlation:*

$$
\begin{aligned}
&\mathbb{E}\left[\left(\sum_{v\in\Gamma_t(u)} \frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2\right] \\
&= \mathbb{E}\left[\sum_{v\in\Gamma_t(u)} \left(\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2 + \sum_{\substack{v,v'\in\Gamma_t(u)\\ v\neq v'}} \frac{(x_t(v)-y_t(v))(x_{v'}^t-y_{v'}^t)}{(1+Z_v)(1+Z_{v'})}\right] \\
&\leq \mathbb{E}\left[|\Gamma_t(u)| \,\middle|\, |\Gamma_t(u)|\geq 1\right] \mathbb{E}\left[\left(\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2\right] \\
&\quad + \mathbb{E}\left[|\Gamma_t(u)|^2 - |\Gamma_t(u)| \,\middle|\, |\Gamma_t(u)|\geq 1\right] \mathbb{E}\left[\frac{x_t(v)-y_t(v)}{1+Z_v}\right]^2.
\end{aligned}
$$

Proof. Our proof strategy is to use conditions that the number of edges is fixed at a certain value.

$$\mathbb{E}\left[\left(\sum_{v\in\Gamma_t(u)}\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2\right]$$
$$=\mathbb{E}\left[\sum_{v\in\Gamma_t(u)}\left(\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2+\sum_{\substack{v,v'\in\Gamma_t(u)\\ v\neq v'}}\frac{(x_t(v)-y_t(v))(x^t_{v'}-y^t_{v'})}{(1+Z_v)(1+Z_{v'})}\right]$$
$$=\mathbb{E}\left[|\Gamma_t(u)|\,\middle|\,|\Gamma_t(u)|\geq 1\right]\mathbb{E}_{Z_v}\left[\left(\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2\right]+\sum_{\substack{v,v'\in\Gamma_t(u)\\ v\neq v'}}\mathbb{E}\left[\frac{(x_t(v)-y_t(v))(x^t_{v'}-y^t_{v'})}{(1+Z_v)(1+Z_{v'})}\right]$$

by the proof below, here we have the inequality

$$\leq\mathbb{E}\left[|\Gamma_t(u)|\,\middle|\,|\Gamma_t(u)|\geq 1\right]\mathbb{E}\left[\left(\frac{x_t(v)-y_t(v)}{1+Z_v}\right)^2\right]$$
$$+\mathbb{E}\left[|\Gamma_t(u)|^2-|\Gamma_t(u)|\,\middle|\,|\Gamma_t(u)|\geq 1\right]\mathbb{E}\left[\frac{x_t(v)-y_t(v)}{1+Z_v}\right]^2.$$

Note that the second equality is true because we can apply Wald's equation on both terms. The number of the first quadratic term is simply $|\Gamma_t(u)|$, as for the second quadratic term, we are choosing 2 different vertices out of $\Gamma_t(u)$.

We now can show the last inequality above:

$$\mathbb{E}\left[\frac{(x_t(v)-y_t(v))(x^t_{v'}-y^t_{v'})}{(1+Z_v)(1+Z_{v'})}\right]$$
$$=\mathbb{E}\left[\frac{(x_t(v)-y_t(v))(x^t_{v'}-y^t_{v'})}{(1+Z_v)(1+Z_{v'})}\,\middle|\,v\sim v'\right]\mathbb{P}\left[v\sim v'\right]$$
$$+\mathbb{E}\left[\frac{(x_t(v)-y_t(v))(x^t_{v'}-y^t_{v'})}{(1+Z_v)(1+Z_{v'})}\,\middle|\,v\nsim v'\right]\mathbb{P}\left[v\nsim v'\right]$$
$$=\mathbb{P}\left[v\sim v'\right]\cdot\sum_{i=0}^{n-3}\sum_{j=0}^{n-3}\frac{(x_t(v)-y_t(v))(x^t_{v'}-y^t_{v'})}{(2+i)(2+j)}\mathbb{P}\left[Z_v=i+1|v\sim v'\right]\mathbb{P}\left[Z_{v'}=j+1|v\sim v'\right]$$
$$+\mathbb{P}\left[v\nsim v'\right]\cdot\sum_{i=0}^{n-3}\sum_{j=0}^{n-3}\frac{(x_t(v)-y_t(v))(x^t_{v'}-y^t_{v'})}{(1+i)(1+j)}\mathbb{P}\left[Z_v=i|v\nsim v'\right]\mathbb{P}\left[Z_{v'}=j|v\nsim v'\right]$$
$$=\mathbb{P}\left[v\sim v'\right]\cdot\sum_{i=0}^{n-3}\frac{x_t(v)-y_t(v)}{(2+i)}\mathbb{P}\left[Z_v=i+1|v\sim v'\right]\sum_{j=0}^{n-3}\frac{x^t_{v'}-y^t_{v'}}{(2+j)}\mathbb{P}\left[Z_{v'}=j+1|v\sim v'\right]$$
$$+\mathbb{P}\left[v\nsim v'\right]\cdot\sum_{i=0}^{n-3}\frac{x_t(v)-y_t(v)}{(1+i)}\mathbb{P}\left[Z_v=i|v\nsim v'\right]\sum_{j=0}^{n-3}\frac{x^t_{v'}-y^t_{v'}}{(1+j)}\mathbb{P}\left[Z_{v'}=j|v\nsim v'\right]$$

$$= \sum_{i=1}^{n-2} \frac{x_t(v) - y_t(v)}{(1+i)} \mathbb{P}\left[Z_v = i, v \sim v'\right] \sum_{j=1}^{n-2} \frac{x_{v'}^t - y_{v'}^t}{(1+j)} \mathbb{P}\left[Z_{v'} = j, v \sim v'\right]$$
$$+ \sum_{i=0}^{n-3} \frac{x_t(v) - y_t(v)}{(1+i)} \mathbb{P}\left[Z_v = i, v \nsim v'\right] \sum_{j=0}^{n-3} \frac{x_{v'}^t - y_{v'}^t}{(1+j)} \mathbb{P}\left[Z_{v'} = j, v \nsim v'\right]$$

Add some zero terms (since the probability is 0) to enable simplification

$$= \sum_{i=0}^{n-2} \frac{x_t(v) - y_t(v)}{(1+i)} \mathbb{P}\left[Z_v = i, v \sim v'\right] \sum_{j=0}^{n-2} \frac{x_{v'}^t - y_{v'}^t}{(1+j)} \mathbb{P}\left[Z_{v'} = j, v \sim v'\right]$$
$$+ \sum_{i=0}^{n-2} \frac{x_t(v) - y_t(v)}{(1+i)} \mathbb{P}\left[Z_v = i, v \nsim v'\right] \sum_{j=0}^{n-2} \frac{x_{v'}^t - y_{v'}^t}{(1+j)} \mathbb{P}\left[Z_{v'} = j, v \nsim v'\right]$$
$$= \mathbb{E}\left[\frac{x_t(v) - y_t(v)}{1+Z_v}, v \sim v'\right] \mathbb{E}\left[\frac{x_{v'}^t - y_{v'}^t}{1+Z_{v'}}, v \sim v'\right]$$
$$+ \mathbb{E}\left[\frac{x_t(v) - y_t(v)}{1+Z_v}, v \nsim v'\right] \mathbb{E}\left[\frac{x_{v'}^t - y_{v'}^t}{1+Z_{v'}}, v \nsim v'\right]$$
$$\leq \left(\mathbb{E}\left[\frac{x_t(v) - y_t(v)}{1+Z_v}, v \sim v'\right] + \mathbb{E}\left[\frac{x_t(v) - y_t(v)}{1+Z_v}, v \nsim v'\right]\right)$$
$$\cdot \left(\mathbb{E}\left[\frac{x_{v'}^t - y_{v'}^t}{1+Z_{v'}}, v \sim v'\right] + \mathbb{E}\left[\frac{x_{v'}^t - y_{v'}^t}{1+Z_{v'}}, v \nsim v'\right]\right)$$
$$= \mathbb{E}\left[\frac{x_t(v) - y_t(v)}{1+Z_v}\right] \mathbb{E}\left[\frac{x_{v'}^t - y_{v'}^t}{1+Z_{v'}}\right]$$

Note that $\mathbb{P}\left[Z_v = 0, v \sim v'\right] = 0, \mathbb{P}\left[Z_v = n-2, v \nsim v'\right] = 0$.

□

5.4.4.2 Calculating (5.4), (5.5)

There is an important quantity in the previous proofs: $\mathbb{E}\left[\frac{1}{1+Z_v}\right]$, which was labeled as (5.4). Now we show how to compute it.

Proposition 5.4.2. *For a Binomial random variable $X \sim \mathsf{Bin}(n,p)$,*

$$\mathbb{E}\left[X | X \geq 1\right] = \frac{np}{1-(1-p)^n}$$

and

$$\mathbb{E}\left[X^2 | X \geq 1\right] = \frac{np(1-p) + n^2p^2}{1-(1-p)^n}$$

and

$$\mathbb{E}\left[X^2 - X | X \geq 1\right] = \frac{n^2p^2 - np^2}{1-(1-p)^n}$$

Proof.

$$\begin{aligned}\mathbb{E}\left[X|X\geq 1\right] &= \sum_{i=0}^{n} i\cdot\mathbb{P}\left[X=i\,|\,X\geq 1\right]\\ &= 0+\sum_{i=1}^{n} i\cdot\frac{\mathbb{P}\left[X=i\right]}{\mathbb{P}\left[X\geq 1\right]} = \frac{\mathbb{E}\left[X\right]}{1-\mathbb{P}\left[X=0\right]} = \frac{np}{1-(1-p)^n}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left[X^2|X\geq 1\right] &= \sum_{i=0}^{n} i^2\cdot\mathbb{P}\left[X=i\,|\,X\geq 1\right]\\ &= 0+\sum_{i=1}^{n} i^2\cdot\frac{\mathbb{P}\left[X=i\right]}{\mathbb{P}\left[X\geq 1\right]} = \frac{\mathbb{E}\left[X^2\right]}{1-\mathbb{P}\left[X=0\right]} = \frac{np(1-p)+n^2p^2}{1-(1-p)^n}\end{aligned}$$

□

Proposition 5.4.3. *For (5.4):* $\mathbb{E}\left[\frac{1}{1+Z_v}\right]$*, we have*

$$\mathbb{E}\left[\frac{1}{1+Z_v}\right]\leq\frac{1-e^{-d}}{d}.$$

Before we prove this proposition, we show the following proposition to help us.

Proposition 5.4.4. *Let* $\mathbf{P}$ *be the transition matrix of a simple random walk on a graph* $G=(V,E)\sim\mathcal{G}(n,p)$*. Then, for any* $u,v\in V$*, it holds that*

$$\mathbb{E}\left[\mathbf{P}(u,v)\right]=\begin{cases}(1-p)^{n-1} & \text{conditioning on } u=v\\ \frac{1-(1-p)^{n-1}}{n-1} & \text{conditioning on } u\neq v\end{cases}$$

Proof. First notice that, for any $u\in V$,

$$\mathbb{E}\left[\mathbf{P}(u,u)\right]=\mathbb{P}\left[u\text{ is isolated}\right]=(1-p)^{n-1}.$$

By symmetry, $\mathbb{E}\left[\mathbf{P}(u,v)\right]=\mathbb{E}\left[\mathbf{P}(u,w)\right]$ for any $v,w\in V\setminus\{u\}$. Moreover, $\mathbb{E}\left[\mathbf{P}\right]$ is a linear combination of stochastic matrices and, hence, stochastic itself. Therefore, for any $u\neq v$,

$$\mathbb{E}\left[\mathbf{P}(u,v)\right]=\frac{1-(1-p)^{n-1}}{n-1}.$$

□

Now we are ready to prove Proposition 5.4.3.

Proof of Proposition 5.4.3. Let $Z_v \sim \mathsf{Bin}(n-2,p)$. Notice that, for $u \neq v$,

$$\mathbb{E}\left[\mathbf{P}(u,v) \mid \{u,v\} \in E\right] = \mathbb{E}\left[\frac{1}{1+Z_v}\right].$$

Since $\mathbb{E}\left[\mathbf{P}(u,v)\right] = \mathbb{E}\left[\mathbf{P}(u,v) \mid \{u,v\} \in E\right] \cdot p$,

$$\mathbb{E}\left[\frac{1}{1+Z_v}\right] = \frac{\mathbb{E}\left[\mathbf{P}(u,v)\right]}{p} = \frac{1-(1-p)^{n-1}}{p(n-1)} \leq \frac{1-e^{-d}}{d}.$$

□

By using a similar trick, we can also show that

$$(5.5): \mathbb{E}\left[\left(\frac{1}{1+Z_v}\right)^2\right] = \frac{2-2e^{-d}d-2e^{-d}}{d^2}.$$

5.5 Missing proofs in Section 5.2

Lemma 5.5.1. *For $\mathcal{G}(n, O(1/n))$ and any constant $\delta > 0$,*

$$\mathbb{P}\left[|\{x \in V : \deg(x) \geq \log\log n\}| \geq n\frac{\delta \log\log^2 n}{\log^2 n}\right] = O\left(\frac{\log^4 n}{n}\right).$$

Proof. The probability that a node has degree greater than k is

$$\mathbb{P}\left[\deg(x) \geq k\right] \leq \binom{n-1}{k}\left(\frac{1}{n}\right)^k \leq \left(\frac{en}{k}\right)^k \left(\frac{1}{n}\right)^k \leq \left(\frac{e}{k}\right)^k.$$

We pick $k = \log\log n$, then let $\gamma = \left(\frac{e}{k}\right)^k = \frac{\log n}{(\log n)^{\log\log\log n}}$. We apply Chebyshev's inequality. Thus need to find the covariance of $\mathbb{1}_{\deg(x)\geq k}$ and $\mathbb{1}_{\deg(y)\geq k}$ for any two nodes x and y. We define a random variable $X = |\{x \in V : \deg(x) \geq \log\log n\}|$.

$$\begin{aligned}
\mathrm{Cov}\left[\mathbb{1}_{\deg(x)\geq k}, \mathbb{1}_{\deg(y)\geq k}\right] &= \mathbb{E}\left[(\mathbb{1}_{\deg(x)\geq k} - \gamma)(\mathbb{1}_{\deg(y)\geq k} - \gamma)\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\deg(x)\geq k}\mathbb{1}_{\deg(y)\geq k}\right] - \gamma^2 \\
&= \mathbb{P}\left[\mathbb{1}_{\deg(x)\geq k}\mathbb{1}_{\deg(y)\geq k}\right] - \gamma^2 \\
&\leq \frac{1}{n}\left(\binom{n-2}{k-1}\left(\frac{1}{n}\right)^{k-1}\right)^2 + \left(1-\frac{1}{n}\right)\left(\binom{n-2}{k}\left(\frac{1}{n}\right)^k\right)^2 - \gamma^2 \\
&\leq \left(\frac{k^2}{e^2 n}\right)\gamma^2 + \left(1-\frac{1}{n}\right)\gamma^2 - \gamma^2 \\
&\leq \gamma^2(1+O(1/n)) - \gamma^2 \\
&\leq O(\gamma^2/n).
\end{aligned}$$

Hence

$$\begin{aligned}\mathrm{Var}\left[X\right] &= n\mathrm{Var}\left[\mathbb{1}_{x\geq k}\right] + \sum_{x,y}\mathrm{Cov}\left[\mathbb{1}_{\deg(x)\geq k}, \mathbb{1}_{\deg(y)\geq k}\right] \\ &\leq n(\gamma-\gamma^2) + n^2 O(\gamma^2/n) \\ &\leq n\gamma - n\gamma^2 + O(n) \\ &= O(n),\end{aligned}$$

and

$$\mathbb{E}\left[X\right] = n\cdot\frac{\log n}{(\log n)^{\log\log\log n}} \ll n\frac{\delta\log\log^2 n}{\log^2 n}.$$

Then

$$\begin{aligned}\mathbb{P}\left[X\geq n\frac{\delta\log\log^2 n}{\log^2 n}\right] &= \mathbb{P}\left[|X-\mathbb{E}\left[X\right]|\geq n\frac{\delta\log\log^2 n}{\log^2 n} - \mathbb{E}\left[X\right]\right] \\ &\leq \frac{\mathrm{Var}\left[X\right]}{(\delta n\log\log^2 n/\log^2 n - \mathbb{E}\left[X\right])^2} \\ &\leq O\left(\frac{\log^4 n}{n}\right),\end{aligned}$$

as required. □

Chapter 6

Conclusion

In this final chapter, we give a conclusion of the thesis and discuss our results and their relevance to previous works again. We will emphasise again the origin and main motivation of the thesis. We also discuss our results and models on a high level, and point to some potential refinements and extensions. In addition, we also explain how our works fit in the related literature and a wider computer science world. At the end of this chapter, we will also conclude potential further research directions.

6.1 Summary of our results

The main motivation of this thesis is to understand how randomisation impacts two classical distributed computing problems: load balancing and random walks on graphs. The two problems are naturally linked by their analyses based on Markov chain theory due to their mathematical nature. Our motivation on adding in the randomness is to make the classical models more realistic, i.e., getting closer to real-world settings. The thesis studies a number of questions regarding the convergence properties of these processes.

In Chapter 1, we present a general picture of the literature and why we are interested in our problems. We also give an introduction of our main contribution and results. Chapter 2 provides sufficient background knowledge to understand the mathematical concepts mainly used in this thesis. Chapter 3 studies the average-case load balancing on networks. In addition, a more comprehensive literature review is given. The novelty of our results is to characterise the time used to reach a small discrepancy when the initial loads are drawn from certain probability distributions instead of considering the worst-case regime set by an adversary. Chapter 4 is concerned with the analysis of random walks on evolving graphs. The edge-Markovian model has been studied in the literature before and is an established and well studied model for dynamic graphs. The contribution of our results is that we do not assume a uniform (i.e., time-invariant) stationary distribution time as in [11, 86]. Only a few works in the literature [80, 89] have done the same. Chapter 5

presents additional results which refine and complement some of the results in Chapter 4. We believe that these results might be of interest in order to analyse other random walk parameters, besides the classical notion of mixing time.

6.2 Significance of our results and models

6.2.1 Average case load balancing

The load balancing problem has been intensively studied for several years [20, 42, 91, 3, 34, 75, 66, 13, 4, 17]. Different aspects and models like diffusion vs. matching, continuous vs. discrete and deterministic vs. randomised, have been intensively studied (see Chapter 3 for more details). They have been used to formalise many problems and usually led to fairly tight convergence bounds. The essence of load balancing contains only these key parts: loads, processors, underlying graphs and how the processors transfer their loads. Since the load balancing protocols are preferred to be local, it is natural to either communicate with *all* neighbours (diffusion) or only one (matching); whether the load can be arbitrarily divided (continuous) or not (discrete); whether we follow exact rules (deterministic) or not (randomised).

Hence we can see that the existing load balancing literature has already established a variety of significant results in different models. The novelty of our approach is, however, to add the randomness somewhere else. Previous randomised load balancing works involve randomisation only in either the rounding of the loads (analysing the error term) or the random matching (relating to underlying graphs). Another important direction in load balancing is called *selfish load balancing* [14, 16, 2]. There, in each round, every token chooses a neighbour randomly and decides probabilistically whether or not to migrate to that neighbour. It is a model which is even more decentralised, because here it is not the nodes who decide with their neighbours how many tokens should be transferred, but that decision is left to the tokens themselves, which are selfish agents.

We analyse a setting with *randomised* initial loads. With this new assumption, we are initiating a new direction apart from the classical worst-case setting. We derive almost matching upper and lower bounds on the discrepancy in this case. More importantly, it also allows us to derive a stronger connection between the time run by the protocol and the achieved discrepancy. We also compared our results and the previous ones in different graph topologies, finding that our case converges faster than the worst-case results.

As an important tool to establish our results, we derived a general mathematical theorem (Theorem 3.5.3), which shows the gap between the t-step transition probability and the corresponding entry in the stationary distribution is bounded by $O\left(\frac{1}{\sqrt{t}}\right)$. Bounding return probabilities of random walks and Markov chains is useful for many other random processes, e.g., [70].

6.2.2 Random walks on evolving graphs

Our study follows a line of previous works of random walks on dynamically changing graphs, e.g., [11, 86, 26, 25, 89]. Dynamic graphs is not new in the random walks field. Evolving graphs, as an extension, emphasise more on the dependencies between two adjacent graphs in the graph sequence. They are often hard to analyse due to their complicated nature. In most research works, some conditions have to be assumed in order to make analysing these problems easier. Therefore, one thing that we overcome in this thesis is that we remove a condition that previous works favour. This makes the problem more realistic but also much harder.

In our study, we study the behaviours of the random walk on a well-studied evolving graph model: the edge-Markovian model. In the graph sequence, each graph (except for the first one) depends on its previous graph. If an edge exists in the previous one, then it has probability q to disappear in the current graph. If an edge does not exist in the previous one, it has probability p to appear in the current one. We provide various results, both positive and negative, for different regimes of this model. Our main observation is that the random walks on the edge-Markovian model can have a mixing behaviour when the graph parameters (p, q) make the graph change slowly and dense. For other regimes, we show negative results like the walk will never mix or weakly mix. The novelty of our work is that we do not assume the graphs all share the same stationary distribution. We also give a way to define the *mixing time* for the random walks in dynamic graphs, which has not been formally given before.

By randomisation, we bring theoretical models one step closer to real world settings. Our proof techniques can potentially help with future works where we want to extend classical results to evolving graphs. For example, the ideas of studying different regimes, separating the process into different stages and classifying regimes using controlled parameters can be useful.

We also propose how to analyse such complicated problems: we clean them up before we analyse them. In our analysis, when encountering the edge-Markovian model, we divide the model into several smaller regimes. This distinction might be useful in future studies, for example, to refine our bounds on mixing times or analyse other random walk parameters such as hitting times or cover times.

6.3 Future work

Generally, it might be hard to develop new protocols other than the diffusion model or the matching model as discussed in Chapter 3. However, the impact of randomness, as the entire thesis suggests, can be powerful. A possible but potentially challenging problem is to study the load balancing in evolving graphs (a combination of the two topics in

this thesis), where the underlying graph is dynamic. Load balancing in dynamic graphs is relatively new compared with previous works. When dynamics are involved, problem usually becomes quite complex in full generality, which may require further restrictions, leading to different models that could be worth exploring.

With respect to random walks on evolving graphs, in our thesis, the analysis given here on the edge-Markovian model is already quite complex. Nevertheless, the changing graph is still (almost) regular and very close to the well-studied Erdős-Rényi random graph. Hence we can see that analysing evolving graphs, even if there is only a minor modification on a well-studied model, can still be very challenging. One main benefit of choosing this edge-Markovian model is that we can borrow some techniques from the results made for the static Erdős-Rényi model. A huge number of results is known for the connectivity (conductance and eigenvalues), diameter, degree distribution, giant component (if the graph is not connected) etc [57, 23, 32, 39, 51, 18]. Since this model was motivated by other distributed computing algorithms, in particular, information spreading research results [25, 26, 27], we would hope that that our random walks results might be useful to understand information spreading, consensus or other dynamics on Markovian-evolving graphs. One possible line of work is [54] where the authors studied dynamic graphs but with a backbone (an invariant) in each graph.

Bibliography

[1] Abramowitz, M., Stegun, I. A., et al. Handbook of mathematical functions. *Applied mathematics series 55* (1966), 62.

[2] Adolphs, C. P., and Berenbrink, P. Distributed selfish load balancing with weights and speeds. In *Proceedings of the 2012 ACM symposium on Principles of distributed computing* (2012), pp. 135–144.

[3] Aiello, W., Awerbuch, B., Maggs, B. M., and Rao, S. Approximate load balancing on dynamic and asynchronous networks. In *Proc. 25th Symp. Theory of Computing (STOC)* (1993), pp. 632–641.

[4] Akbari, H., and Berenbrink, P. Parallel rotor walks on finite graphs and applications in discrete load balancing. In *Proceedings of the twenty-fifth annual ACM symposium on Parallelism in algorithms and architectures* (2013), pp. 186–195.

[5] Akbari, H., Berenbrink, P., and Sauerwald, T. A simple approach for adapting continuous load balancing processes to discrete settings. In *Proceedings of the 2012 ACM symposium on Principles of distributed computing* (2012), pp. 271–280.

[6] Alistarh, D., Censor-Hillel, K., and Shavit, N. Are lock-free concurrent algorithms practically wait-free? *J. ACM 63*, 4 (2016), 31:1–31:20.

[7] Alon, N., and Milman, V. D. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B 38*, 1 (1985), 73–88.

[8] Alon, N., and Spencer, J. *The Probabilistic Method*, 2nd ed. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, 2000.

[9] Anagnostopoulos, A., Kirsch, A., and Upfal, E. Load balancing in arbitrary network topologies with stochastic adversarial input. *SIAM J. Comput. 34*, 3 (2005), 616–639.

[10] Augustine, J., Pandurangan, G., and Robinson, P. Distributed algorithmic foundations of dynamic networks. *SIGACT News 47*, 1 (2016), 69–98.

[11] Avin, C., Koucký, M., and Lotker, Z. Cover time and mixing time of random walks on dynamic graphs. *Random Structures Algorithms 52*, 4 (2018), 576–596.

[12] Becchetti, L., Clementi, A. E. F., Natale, E., Pasquale, F., and Trevisan, L. Find your place: Simple distributed algorithms for community detection. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'17)* (2017), pp. 940–959.

[13] Berenbrink, P., Cooper, C., Friedetzky, T., Friedrich, T., and Sauerwald, T. Randomized diffusion for indivisible loads. *Journal of Computer and System Sciences 81*, 1 (2015), 159–185.

[14] Berenbrink, P., Friedetzky, T., Goldberg, L. A., Goldberg, P. W., Hu, Z., and Martin, R. Distributed selfish load balancing. *SIAM Journal on Computing 37*, 4 (2007), 1163–1181.

[15] Berenbrink, P., Giakkoupis, G., Kermarrec, A.-M., and Mallmann-Trenn, F. Bounds on the voter model in dynamic networks. In *Proceedings of the 43rd International Colloquium on Automata, Languages and Programming (ICALP)* (July 12–15 2016), pp. 146:1–146:15.

[16] Berenbrink, P., Hoefer, M., and Sauerwald, T. Distributed selfish load balancing on networks. *ACM Transactions on Algorithms (TALG) 11*, 1 (2014), 1–29.

[17] Berenbrink, P., Klasing, R., Kosowski, A., Mallmann-Trenn, F., and Uznański, P. Improved analysis of deterministic load-balancing schemes. *ACM Transactions on Algorithms (TALG) 15*, 1 (2018), 1–22.

[18] Berestycki, N., Lubetzky, E., Peres, Y., Sly, A., et al. Random walks on the random graph. *The Annals of Probability 46*, 1 (2018), 456–490.

[19] Berry, A. C. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society 49*, 1 (1941), 122–136.

[20] Boillat, J. E. Load balancing and poisson equation in a graph. *Concurrency: Pract. Exper. 2* (1990), 289–313.

[21] Boyd, S., Ghosh, A., Prabhakar, B., and Shah, D. Gossip and mixing times of random walks on random graphs, 2004.

[22] Boyd, S., Ghosh, A., Prabhakar, B., and Shah, D. Randomized Gossip Algorithms. *IEEE Transactions on Information Theory and IEEE/ACM Transactions on Networking 52* (2006), 2508–2530.

[23] Boyd, S. P., Ghosh, A., Prabhakar, B., and Shah, D. Mixing times for random walks on geometric random graphs. In *ALENEX/ANALCO* (2005), pp. 240–249.

[24] Chung, F., and Zhao, W. PageRank and random walks on graphs. *Bolyai Society Mathematical Studies 20* (2010), 43–62.

[25] Clementi, A., Crescenzi, P., Doerr, C., Fraigniaud, P., Pasquale, F., and Silvestri, R. Rumor spreading in random evolving graphs. *Random Structures and Algorithms 48*, 2 (2016), 290–312.

[26] Clementi, A., Monti, A., Pasquale, F., and Silvestri, R. Information spreading in stationary Markovian evolving graphs. *IEEE Transactions on Parallel and Distributed Systems 22*, 9 (2011), 1425–1432.

[27] Clementi, A. E. F., Macci, C., Monti, A., Pasquale, F., and Silvestri, R. Flooding time of edge-markovian evolving graphs. *SIAM J. Discrete Math. 24*, 4 (2010), 1694–1712.

[28] Clementi, A. E. F., Silvestri, R., and Trevisan, L. Information spreading in dynamic graphs. *Distributed Computing 28*, 1 (2015), 55–73.

[29] Cooper, B. F. Quickly routing searches without having to move content. In *International Workshop on Peer-to-Peer Systems* (2005), Springer, pp. 163–172.

[30] Cooper, C., Elsasser, R., Ono, H., and Radzik, T. Coalescing random walks and voting on connected graphs. 1–16.

[31] Cooper, C., Elsasser, R., Ono, H., and Radzik, T. Coalescing random walks and voting on connected graphs. *SIAM Journal on Discrete Mathematics 27*, 4 (2013), 1748–1758.

[32] Cooper, C., and Frieze, A. The cover time of sparse random graphs. *Random Structures & Algorithms 30*, 1-2 (2007), 1–16.

[33] Cooper, C., Frieze, A., and Radzik, T. Multiple random walks and interacting particle systems. *Automata, Languages and Programming* (2009), 1–28.

[34] Cybenko, G. Dynamic load balancing for distributed memory multiprocessors. *Journal of parallel and distributed computing 7*, 2 (1989), 279–301.

[35] Cybenko, G. Load balancing for distributed memory multiprocessors. *J. Parallel and Distributed Comput. 7* (1989), 279–301.

[36] Denysyuk, O., and Rodrigues, L. E. T. Random walks on evolving graphs with recurring topologies. In *Distributed Computing - 28th International Symposium, DISC 2014, Austin, TX, USA, October 12-15, 2014. Proceedings* (2014), pp. 333–345.

[37] Esseen, C.-G. *On the Liapounoff limit of error in the theory of probability.* Almqvist & Wiksell, 1942.

[38] Fogaras, D., Rácz, B., Csalogány, K., and Sarlós, T. Towards scaling fully personalized pagerank: Algorithms, lower bounds, and experiments. *Internet Mathematics 2*, 3 (2005), 333–358.

[39] Fountoulakis, N., and Reed, B. A. The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Structures & Algorithms 33*, 1 (2008), 68–86.

[40] Friedrich, T., Gairing, M., and Sauerwald, T. Quasirandom load balancing. *SIAM Journal on Computing 41*, 4 (2012), 747–771.

[41] Friedrich, T., and Sauerwald, T. Near-perfect load balancing by randomized rounding. *Proceedings of the 41st annual ACM symposium on Symposium on theory of computing - STOC '09* (2009), 121.

[42] Ghosh, B., Leighton, F. T., Maggs, B. M., Muthukrishnan, S., Plaxton, C. G., Rajaraman, R., Richa, A. W., Tarjan, R. E., and Zuckerman, D. Tight analyses of two local load balancing algorithms. *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing - STOC '95 29*, 1 (1995), 548–558.

[43] Ghosh, B., and Muthukrishnan, S. Dynamic load balancing by random matchings. *Journal of Computer and System Sciences 53*, 3 (1996), 357–370.

[44] Giakkoupis, G., Sauerwald, T., and Stauffer, A. Randomized rumor spreading in dynamic graphs. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part II* (2014), pp. 495–507.

[45] Goel, A., and Indyk, P. Stochastic load balancing and related problems. In *40th Annual Symposium on Foundations of Computer Science (FOCS)* (1999), pp. 579–586.

[46] Goldstein, M. L., Morris, S. A., and Yen, G. G. Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter and Complex Systems 41*, 2 (2004), 255–258.

[47] Hermon, J., and Sousi, P. Random walk on dynamical percolation. *arXiv preprint arXiv:1902.02770* (2019).

[48] Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association 58*, 301 (1963), 13–30.

[49] Hoffman, C., Kahle, M., and Paquette, E. Spectral Gaps of Random Graphs and Applications. *International Mathematics Research Notices* (05 2019).

[50] Huebner, K. H., Dewhirst, D. L., Smith, D. E., and Byrom, T. G. *The Finite Element Methods for Engineers*. Wiley, 2001.

[51] Jonasson, J. On the cover time for random walks on random graphs. *Combinatorics, Probability and Computing 7*, 3 (1998), 265–279.

[52] Karger, D. R., and Ruhl, M. Simple efficient load balancing algorithms for peer-to-peer systems. In *Proceedings of the sixteenth annual ACM symposium on Parallelism in algorithms and architectures* (2004), pp. 36–43.

[53] Kempe, D., Kleinberg, J., and Demers, A. Spatial gossip and resource location protocols. *Journal of the ACM (JACM) 51*, 6 (2004), 943–967.

[54] Kuhn, F., Lynch, N., and Oshman, R. Distributed computation in dynamic networks. In *Proceedings of the forty-second ACM symposium on Theory of computing* (2010), pp. 513–522.

[55] Kuhn, F., and Oshman, R. Dynamic networks: models and algorithms. *SIGACT News 42*, 1 (2011), 82–96.

[56] Lamprou, I., Martin, R., and Spirakis, P. Cover time in edge-uniform stochastically-evolving graphs. *Algorithms (Basel) 11*, 10 (2018), Paper No. 149, 15.

[57] Lee, J. R., Gharan, S. O., and Trevisan, L. Multi-way spectral partitioning and higher-order Cheeger inequalities.

[58] Levin, D. A., and Peres, Y. *Markov chains and mixing times*, vol. 107. American Mathematical Soc., 2017.

[59] Levin, D. A., Peres, Y., and Wilmer, E. L. *Markov chains and mixing times*. American Mathematical Soc., 2009.

[60] Lovász, L., and Winkler, P. Efficient Stopping Rules for Markov Chains (Extended Abstract). *Proc. 27th ACM Symp. on the Theory of Computing* (1995), 76–82.

[61] Lyons, R. Asymptotic enumeration of spanning trees. *Combinatorics, Probability and Computing 14* (2005), 491–522.

[62] Mavronicolas, M., and Sauerwald, T. The impact of randomization in smoothing networks. *Distributed Computing 22*, 5-6 (2010), 381–411.

[63] McDiarmid, C. Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics*, M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, Eds. Springer-Verlag, 1998, pp. 195–243.

[64] Michail, O., and Spirakis, P. G. Elements of the theory of dynamic networks. *Commun. ACM 61*, 2 (2018), 72.

[65] Montenegro, R., and Tetali, P. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci. 1*, 3 (2006), x+121.

[66] Muthukrishnan, S., Ghosh, B., and Schultz, M. H. First- and second-order diffusive methods for rapid, coarse, distributed load balancing. *Theory Comput. Syst. 31* (1998), 331–354.

[67] Muthukrishnan, S., Ghosh, B., and Schultz, M. H. First- and second-order diffusive methods for rapid, coarse, distributed load balancing. *Theory Comput. Syst. 31*, 4 (1998), 331–354.

[68] Newman, M., Barabási, A.-L., and Watts, D. J. *The structure and dynamics of networks*. Princeton university press, 2011.

[69] Norris, J. R. *Markov chains*. No. 2. Cambridge university press, 1998.

[70] Oliveira, R. I., and Peres, Y. Random walks on graphs: new bounds on hitting, meeting, coalescing and returning. In *2019 Proceedings of the Sixteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)* (2019), SIAM, pp. 119–126.

[71] Panconesi, A., and Srinivasan, A. Improved distributed algorithms for coloring and network decomposition problems. In *Proc. 24th Symp. Theory of Computing (STOC)* (1992), pp. 581–592.

[72] Panconesi, A., and Srinivasan, A. Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds. *SIAM Journal on Computing 26*, 2 (1997), 350–368.

[73] Peres, Y., Sousi, P., and Steif, J. Mixing time for random walk on supercritical dynamical percolation. *Prob. Theory and Related Fields* (2019). (to appear).

[74] Peres, Y., Stauffer, A., and Steif, J. E. Random walks on dynamical percolation: mixing times, mean squared displacement and hitting times. *Probab. Theory Related Fields 162*, 3-4 (2015), 487–530.

[75] Rabani, Y., Sinclair, A., and Wanka, R. Local divergence of Markov chains and the analysis of iterative load-balancing schemes. *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No.98CB36280)* (1998), 694–703.

[76] Saloff-Coste, L., and Zúñiga, J. Merging for time inhomogeneous finite Markov chains. I. Singular values and stability. *Electron. J. Probab. 14* (2009), 1456–1494.

[77] Saloff-Coste, L., and Zúñiga, J. Merging and stability for time inhomogeneous finite Markov chains. In *Surveys in stochastic processes*, EMS Ser. Congr. Rep. Eur. Math. Soc., Zürich, 2011, pp. 127–151.

[78] Saloff-Coste, L., and Zúñiga, J. Merging for inhomogeneous finite Markov chains, Part II: Nash and log-Sobolev inequalities. *Ann. Probab. 39*, 3 (2011), 1161–1203.

[79] Saloff-Coste, L., and Zúñiga, J. Merging for time inhomogeneous finite markov chains, part I: Singular values and stability. *Electronic Journal of Probability 14* (2009), 1456–1494.

[80] Saloff-Coste, L., and Zúñiga, J. Merging and stability for time inhomogeneous finite Markov chains. *arXiv preprint arXiv:1004.2296* (2010), 1–26.

[81] Sanders, P. Analysis of nearest neighbor load balancing algorithms for random loads. *Parallel Computing 25*, 8 (1999), 1013–1033.

[82] Sarma, A. D., Gollapudi, S., and Panigrahy, R. Estimating pagerank on graph streams. *Journal of the ACM (JACM) 58*, 3 (2011), 1–19.

[83] Sarma, A. D., Molla, A. R., and Pandurangan, G. Distributed computation in dynamic networks via random walks. *Theor. Comput. Sci. 581* (2015), 45–66.

[84] Sarma, A. D., Nanongkai, D., Pandurangan, G., and Tetali, P. Distributed random walks. *J. ACM 60*, 1 (2013), 2:1–2:31.

[85] Sauerwald, T., and Sun, H. Tight bounds for randomized load balancing on arbitrary network topologies. In *Proc. 53rd Symp. Foundations of Computer Science (FOCS)* (2012), pp. 341–350.

[86] Sauerwald, T., and Zanetti, L. Random walks on dynamic graphs: mixing times, hitting times, and return probabilities. In *46th International Colloquium on Automata, Languages, and Programming*, vol. 132 of *LIPIcs. Leibniz Int. Proc. Inform.* 2019, pp. Art. No. 93, 15.

[87] Sauerwald, T., and Zanetti, L. Random walks on dynamic graphs: Mixing times, hittingtimes, and return probabilities. *arXiv preprint arXiv:1903.01342* (2019).

[88] Sinclair, A., and Jerrum, M. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation 82*, 1 (1989), 93–133.

[89] Sousi, P., and Thomas, S. Cutoff for random walk on dynamical Erdos-Renyi graph. *arXiv preprint arXiv:1807.04719* (2018).

[90] Sousi, P., Thomas, S., et al. Cutoff for random walk on dynamical erdős–rényi graph. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* (2020), vol. 56, Institut Henri Poincaré, pp. 2745–2773.

[91] Subramanian, R., and Scherson, I. D. An analysis of diffusive load-balancing. In *Proc. 6th Symp. Parallelism in Algorithms and Architectures (SPAA)* (1994), pp. 220–225.

[92] Surana, S., Godfrey, B., Lakshminarayanan, K., Karp, R., and Stoica, I. Load balancing in dynamic structured peer-to-peer systems. *Performance Evaluation 63*, 3 (2006), 217–240.

[93] Williams, R. D. Performance of dynamic load balancing algorithms for unstructured mesh calculations. *Concurrency: Practice and Experience 3*, 5 (1991), 457–481.

[94] Zhanga, D., Jianga, C., and Li, S. A fast adaptive load balancing method for parallel particle-based simulations. *Simulation Modelling Practice and Theory 17*, 6 (2009), 1032–1042.